LiftEd
Transforming Foundational Literacy and Numeracy

EDTECH ACCELERATOR

LiftEd EdTech Accelerator

Impact on Learning Outcomes Study

Ei Mindspark

Rajasthan | 2026

Principal Investigator
Prof. Tarun Jain
Professor of Economics, Indian Institute of Management Ahmedabad

This study was conducted by Educational Initiatives under the guidance of
Tarun Jain - Professor of Economics and Reserve Bank of India Chair in Finance &
Economics at the Indian Institute of Management Ahmedabad.

Table of Contents

Acknowledgements

We express our profound appreciation to those instrumental in conducting this study and compiling the report. Their support has been indispensable to this endeavour.

We extend our gratitude to the Principal Investigator, Professor Tarun Jain, Associate Professor of Economics and the Reserve Bank of India Chair in Finance and Economics at the Indian Institute of Management Ahmedabad. His guidance in designing the study, overseeing the creation of robust data collection tools and ensuring the quality of insights has been pivotal.

We express our gratitude to the LiftEd EdTech Accelerator consortium partners - Michael & Susan Dell Foundation, Reliance Foundation, UBS Optimus Foundation and British Asian Trust for their flexibility to innovate and generous support throughout the project. We also extend our gratitude to acknowledge the support and counsel from Central Square Foundation.

We extend our deepest gratitude to the team at Educational Initiatives for their expertise in conducting the study, generating insights, and report writing. We are especially grateful to Sridhar Rajagopalan (Co-Founder), Pranav Kothari (CEO), Gayatri Vaidya (Associate Vice President), Arvind Singh (Associate Vice President), and Ei's Product team led by Sangeeta Oak (Associate Vice President) for their leadership and guidance. We would also like to acknowledge Dipak Chalakkal, Vaishali Tiwari, Prashant Kapoor, and Aparna Sanjay, along with the field Operations team and enumerators who conducted the assessments, whose dedication and efforts were instrumental in making this study a reality.

Finally and most importantly, we are indebted to the participants of the study for their time and valuable contributions and the EdTech portfolio partner: Ei Mindspark for their support in facilitating the planning and coordination of the study.

We hope the study findings will inform stakeholders' decisions on designing and implementing effective EdTech interventions and strengthen the collective evidence on EdTech for foundational learning.

In deep gratitude,
Educational Initiatives

Glossary & Abbreviations

Term	Description
Addition/ Subtraction Process	Refers to tasks that assess students' ability to solve multi-digit or column-based addition and subtraction problems, requiring correct sequencing of procedural steps such as carrying over or borrowing. These tasks go beyond fact recall and capture procedural fluency in formal arithmetic methods (e.g., solving 47 + 38 using column addition).
Attrition Buffer	Allowance built into the study's sample design to account for the expected loss of participants over time. As the study was longitudinal and conducted over several months, some students present at baseline were unavailable at endline due to absence or mobility and therefore an attrition buffer was included to preserve statistical power.
Conceptual Skills[1]	Refers to competencies that require application, reasoning, integration of multiple concepts and problem solving. These skills typically involve multi-step, non-routine, or contextualised tasks and may include spatial, or language-heavy demands, such as solving word problems or applying arithmetic concepts to real-life situations. The tasks from the Numeracy tool classified as conceptual Skills for the purposes of this study are Addition Process (2 digit, column addition), Subtraction Process (2 digit, column addition), Counting in Bundles, Missing Number, Word Problems and Shape Recognition.
Confidence	Refers to the degree of certainty that observed results reflect a true underlying effect rather than chance. It is influenced by sample size, outcome variability, study design and precision of estimates, often reflected through p-values and confidence intervals.
Confidence Intervals (CI)	Represent a statistical range around an estimated effect that conveys the degree of uncertainty in the estimate. They indicate the range within which the true population effect is likely to lie, assuming the underlying model is correct (e.g., an average of estimate of 0.7 year with a 95% confidence interval of 0.6 to 0.9).
Difference-in-Differences (DiD)	Analytical method used to estimate programme impact by comparing changes in outcomes over time between Intervention and Comparison groups. This approach helps account for baseline differences and common time trends affecting both groups.

1 Ministry of Education, Government of India. (2021). National initiative for proficiency in reading with understanding and numeracy (NIPUN Bharat): Mission guidelines. https://www.education.gov.in/sites/upload_files/mhrd/files/nipun_bharat_guidelines.pdf .

Term	Description
Intent-to-Treat (ITT)	Impact estimate that captures the effect of being assigned to the intervention, irrespective of actual participation or usage levels. ITT estimates preserve the original group assignment and provide a conservative measure of programme effectiveness under real-world conditions.
Intra-Class Correlation (ICC)	Refers to the degree to which students within the same class or school resemble one another in terms of learning outcomes, reducing the amount of independent information contributed by each student.
Minimum Detectable Effect Size	The smallest difference between Intervention and Comparison groups that the study is statistically powered to detect, given assumptions about sample size, outcome variability, clustering and significance levels.
Outcome Variability/Variance	Refers to the amount of natural variation in outcomes across individuals rather than uniform responses. Higher variability in student performance increases uncertainty in estimates and typically requires larger samples to detect programme effects.
Power	Probability that a study will correctly detect a true effect of a specified size if it exists in the population. Studies are commonly designed with power levels of 80 percent or higher to reduce the likelihood of false negatives.
Procedural Skills[2]	Refers to tasks that primarily assess recall, recognition and the execution of well-rehearsed procedures, with minimal reasoning demands. These tasks typically focus on basic number sense and fact recall, involve limited interpretation or transfer and are relatively language-light. For the purposes of this study, the following tasks in the numeracy tool are classified as Procedural Skills: Number Comparison, Addition Facts (untimed), Addition Facts (timed), Subtraction Facts (untimed) and Subtraction Facts (timed).
Quasi-Experimental Design (QED)	Study design used to estimate the effects of an intervention when random assignment of participants to treatment and comparison groups is not feasible. Instead, the design relies on non-randomly formed groups and applies statistical or design-based techniques (such as matching or difference-in-differences) to control for pre-existing differences and approximate causal inference.
Standard Deviation (SD)	Statistical measure of variability that indicates how spread out individual observations are around the mean. When used to express effect sizes, SD units standardise differences across groups or time points, allowing impacts to be compared across outcomes measured on different scales.
Treatment-on-the-Treated (ToT)	Impact estimate that reflects the effect of the intervention on participants who actually received or engaged with it as intended. ToT estimates adjust for differential usage or compliance and are typically derived using assignment to treatment as an instrument for actual participation.

2 Ministry of Education, Government of India. (2021). National initiative for proficiency in reading with understanding and numeracy (NIPUN Bharat): Mission guidelines. https://www.education.gov.in/sites/upload_files/mhrd/files/nipun_bharat_guidelines.pdf .

1. Executive Summary

This chapter outlines the purpose, approach and key findings from the independent evaluation of the Ei Mindspark EdTech solution in Rajasthan, examining its effectiveness in improving foundational numeracy outcomes.

1.1 About the Study

Foundational Literacy and Numeracy (FLN) underpin lifelong learning in India, yet significant learning gaps emerge early and persist despite recent strides under NEP 2020, NIPUN Bharat Mission and increased FLN investments. Inclusive EdTech presents a strong opportunity to strengthen FLN at home and in schools, supported by growing smartphone access. However, most existing solutions remain poorly aligned with the languages, needs and contexts of low-income communities.

The LiftEd EdTech Accelerator was created to bridge this gap by supporting contextually relevant, pedagogically sound EdTech for “Low-Income Bharat.” Working with governments and ecosystem partners, it enables large-scale adoption through funding, mentorship and technical assistance for eight leading EdTech organisations. A robust evaluation agenda focused on learning outcomes, impact and user experience has generated actionable insights. Collectively, LiftEd-supported solutions have reached over 5.5 million users, directly benefiting more than 3 million children in improving foundational learning.

To build evidence on EdTech’s effectiveness in FLN, the Accelerator commissioned an independent evaluation to identify what works, its impact on learning outcomes and key enablers for scale. Conducted by Educational Initiatives (Ei) with Central Square Foundation, the study assessed selected solutions’ impact on early mathematics across grades and geographies. As the outcome evaluator, Ei consolidates findings at the partner level, measuring the impact on learning outcomes as a result of the solution usage.

The EdTech solution evaluated in Rajasthan is Ei Mindspark, a personalised adaptive platform that creates customised learning pathways in Mathematics through a simple, gamified interface.

1.2 Study Design

The study assessed the impact of Ei Mindspark’s FLN intervention on numeracy outcomes for Grades 2 and 3 students in Ajmer, Bundi and Kota districts in Rajasthan using a longitudinal quasi-experimental design. The study uses a difference-in-differences approach[3] comparing learning growth between intervention and comparison groups over time, with the resulting impact estimates subsequently converted into Equivalent Years of Schooling (EYOS) to aid interpretation. The Baseline round was conducted in August - October 2024 and the Endline round in March - April 2025. Assessments were administered through the Ei NEEV application, capturing student responses via audio and touch-based inputs.

3 The DiD effect size is calculated as: [Avg Delta_Intervention (Δi) - Avg. Delta_Comparison (Δc)] / Pooled Standard Deviation]

1.3 Key Findings

1.3.1 The Ei Mindspark Programme delivered strong gains in numeracy, with significant impact across number sense and arithmetic fluency, as well as problem-solving.

The evaluation demonstrates that the programme achieved statistically significant learning gains across a wide range of numeracy competencies. The strongest effects were observed in Subtraction Facts timed (0.81 SD), Subtraction Process (0.78 SD), Addition Facts timed (0.69 SD), Addition Process (0.69 SD) and Shape Recognition (0.71 SD). These results highlight the programme's effectiveness in strengthening foundational arithmetic fluency as well as enabling deeper improvements in conceptual understanding and applied problem-solving.

The EYOS analysis indicates meaningful learning acceleration of ~1.6 times relative to Business-As-Usual instruction. The intervention generated learning gains equivalent to a substantial fraction of a year of schooling over the study period, demonstrating that students progressed faster than peers in comparison schools. These findings suggest that the model not only improves absolute learning levels but also meaningfully compresses the time required to achieve expected grade-level competencies, highlighting its potential to address foundational learning gaps in low-performing contexts.

1.3.2 Both Grades recorded strong numeracy gains, with Grade 2 showing broad foundational improvements and Grade 3 excelling in conceptual tasks.

Students in both Grades 2 and 3 demonstrated notable improvements, though with different areas of relative strength. Grade 2 achieved the strongest results in Subtraction Process (0.93 SD), Missing Number (0.82 SD) and Word Problems (0.70 SD), indicating wide-ranging progress across numeracy tasks. Grade 3 also recorded significant gains, particularly in Subtraction Process (0.68 SD), Shape Recognition (0.70 SD) and Addition Process (0.62 SD). Overall, the programme effectively consolidated early arithmetic skills for lower grade learners while supporting the higher grade students in advancing to conceptual numeracy competencies.

1.3.3 Girls recorded stronger gains in fluency-oriented procedural tasks, while boys showed larger improvements in conceptual numeracy and problem-solving.

The programme improved numeracy outcomes for both boys and girls, with patterns differentiated by gender. Girls' largest gains appeared in timed arithmetic fluency tasks such as Subtraction Facts (0.87 SD) and Addition Facts (0.73 SD). Boys recorded broader gains across untimed and conceptual tasks, including Subtraction Process (0.86 SD) and Word Problems (0.64 SD), demonstrating relatively stronger growth in conceptual reasoning and problem-solving.

1.3.4 The intervention enabled stronger upward mobility from lower performance levels in numeracy and supported stability and progress at higher levels.

Intervention students were significantly less likely to remain at the lowest level (L0)[4] and substantially more likely to progress to higher proficiency levels than those in the Comparison group. For example, 91% of Intervention students who were at L0 in Addition Facts advanced to L3 or L4, compared to 71% of Comparison students. Mid-level students (L2) also demonstrated stronger upward mobility under the Intervention, with a greater share progressing from L2 to higher levels (L3/L4) than their Comparison counterparts. At the higher end of the proficiency distribution, students under the Intervention exhibited better stability and progression—for instance, 43% of students at L4 in Word Problems maintained their level at Endline. These findings underscore consistent upward mobility among low- and mid-performing students, alongside relatively greater stability for high performers.

The Ei Mindspark Programme demonstrated strong and consistent impact on numeracy outcomes, with particularly large gains in arithmetic fluency, multi-step processes, and problem-solving. Effects were visible across grades, gender, and performance levels, with distinct patterns of strength across subgroups. Higher engagement further amplified learning, underscoring the importance of sustained usage for advanced numeracy development.

4 Categories L0–L4 classify student performance from zero scores to highest proficiency, based on accuracy percentages and fluency relative to the baseline average. These levels help track patterns of improvement or decline across the evaluation timeline. L0 = Zero scorers, L1 = 0 to 25% (accuracy tasks) and 0 to 0.25x Baseline Avg (fluency tasks). Further details on the classification are presented in Section 4.4.

2. Background of the Study

2.1 Inclusive EdTech for Foundational Learning

Foundational Literacy and Numeracy (FLN) - the ability to read, write, comprehend and perform basic mathematical operations by the end of Grade 3 - is widely recognised as the cornerstone of lifelong learning.[5, 6, 7, 8] In India, significant learning deficits emerge early, particularly among children aged 4 to 8, with large proportions of students struggling to acquire basic skills. Repeated rounds of the Annual Status of Education Report (ASER) over the past decade have consistently highlighted low levels of foundational learning. These early gaps make it increasingly difficult for children to grasp more complex concepts as they move through higher grades. The most recent ASER data, however, points to green shoots of progress - improvements in basic reading levels for Grade 3 government students being the highest since 2005 with basic arithmetic levels also showing substantial improvement.[9] Yet, the depth and persistence of learning deficits across the system make clear that far more remains to be done.

To address these challenges, the Government of India has launched multiple initiatives to strengthen FLN, including NEP 2020, which frames FLN as a five-year continuum and emphasises curriculum reform, technology integration and teacher capacity-building. NIPUN Bharat (2021) sets FLN targets for preschool to Grade 3 by 2026–27, supported by initiatives such as the CBSE Reading Mission, the National Curriculum Framework for the Foundational Stage and Jaadui Pitaara, which together provide age-appropriate, multilingual learning resources. Progress is monitored through large-scale assessments like NAS and FLS, while increased funding under Samagra Shiksha and state-level programmes, often implemented with private and civil society partners, aim to further bridge learning gaps[10,11] Further, budget allocations for education under Samagra Shiksha have increased, from ₹31,050 crores (~4200 million USD) in FY 2021-22 to ₹41,250 crore in FY 2025-26.[12] Concurrent to the central government initiatives, state governments, in collaboration with private partners and civil society organisations, have implemented tailored initiatives to bridge learning gaps.

In this context, educational technology (EdTech) has emerged as a potential multiplier, harnessing digital tools to support teaching and learning both in classrooms and at home. A recurring question, however, is whether EdTech is truly accessible to all, especially children from low-income households. Emerging data suggests that access to digital

5 Ministry of Human Resource Development, Government of India. (2020). *National Education Policy 2020.* https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
6 Institute for Competitiveness. (n.d.). *State of Foundational Literacy and Numeracy in India.* https://www.competitiveness.in/wp-content/uploads/2021/12/Report_on_state_of_foundational_learning_and_numeracy_web_version.pdf
7 Ministry of Education, Government of India. (n.d.). *About Foundation Literacy and Numeracy.* https://diksha.gov.in/fln.html.
8 Sinha, A (2023). Maximising India's demographic dividend through foundational literacy and numeracy. *Hindustan Times.*(https://www.hindustantimes.com/ht-insight/knowledge/maximising-india-s-demographic-dividend-through-foundational-literacy-and-numeracy-101699332886168.html).
9 Annual Status of Education Report (ASER). (2024). *ASER 2024 National Findings* https://asercentre.org/wp-content/uploads/2022/12/ASER-2024-National-findings.pdf
10 Ministry of Human Resource Development, Government of India. (2020). *National Education Policy 2020.* https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
11 Storyweaver (n.d.). *CBSE Reading Mission.* https://storyweaver.org.in/en/about/campaigns/cbse-reading-mission
12 CNBC TV 18. (2025). Budget 2025: *National Education Mission receives outlay of ₹41,250 crore* https://www.cnbctv18.com/budget/budget-2025-national-education-mission-samagra-shiksha-abhiyaan-receives-outlay-of-rs-41250-crore-19548980.htm

infrastructure in these contexts is becoming increasingly feasible, particularly through smartphones. Studies indicate that **90% of households** have access to at least one smartphone and in **75% of these households**, children regularly use the device - typically spending around an hour on it each day.[13] This rise in smartphone penetration, coupled with growing internet availability, presents a significant opportunity to leverage EdTech solutions to improve learning outcomes and narrow educational gaps.

Building on this opportunity, EdTech offers innovative solutions to address critical educational challenges, such as varying teacher quality, diverse learning levels within classrooms and limited access to quality instructional resources. By equipping teachers with tools for effective pedagogy and enabling parents to support their children through interactive content and progress tracking, well-designed, pedagogically sound EdTech solutions have the potential to significantly improve learning outcomes.

Globally, there is emerging evidence on the potential of EdTech to support learning at home. The Global Learning XPRIZE Competition, launched in 2014, incentivised teams from around the world to create open-sourced, scalable software that empowers children to achieve foundational learning skills and saw learning gains for both literacy and numeracy across competing solutions.[14] Similarly, Angrist, Bergman and Matsheng provide experimental evidence on strategies to support learning when schools close.[15] Using a randomised control design, they tested two low-technology interventions in Botswana – SMS messages and phone calls – with parents to support their child's learning and found that combined treatment improves learning by 0.12 standard deviations. This translates to 0.89 standard deviations of learning per USD 100, ranking among the most cost-effective interventions to improve learning.

Despite this promise, most current EdTech solutions in India are designed primarily for middle- and high-income users. Content is often in English, misaligned with the lived realities and languages of children from low-income communities and offered at price points that place it beyond their reach. As a result, the children who could benefit most from effective EdTech are often the least likely to access it. Unlocking EdTech's transformative potential for "Low Income Bharat" therefore requires inclusive, affordable solutions that are contextually relevant, explicitly address foundational learning and are backed by rigorous evidence of effectiveness. Bridging these gaps will ensure that EdTech becomes a critical lever for equitable and impactful education in India. It was with this objective that the **LiftEd EdTech Accelerator** was set up.

2.2 LiftEd EdTech Accelerator

To bridge the gaps as defined above and to leverage the opportunity that India has, a consortium of non-profit and philanthropic organisations have set up a LiftEd EdTech Accelerator, a two-year initiative from April 2023-25, to support foundational learning of children using EdTech. The Accelerator was set up to support the NIPUN Bharat mission to significantly shape the future of tech-based learning at home for foundational literacy and numeracy in India by reaching 2.5 million children by 2025.

13 Central Square Foundation. (2026). *Bharat Survey for EdTech (BaSE) Report 2026.* Bharat Survey for EdTech (BaSE) Report 2025
14 Global Learning X Prize. (n.d.). *Global Learning X Prize: Executive Summary* https://assets-us-01.kc-usercontent.com/5cb25086-82d2-4c89-94f0-8450813a0fd3/fc467c7f-d8bd-4d05-bba3-2aa9b06833fb/GLEXP_Executive%20Summary.pdf
15 Angrist, N., Bergman, P. & Matsheng, M. (2022). Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour*, 6, 941-950. https://doi.org/10.1038/s41562-022-01381-z

The LiftEd EdTech Accelerator is anchored by Michael & Susan Dell Foundation, Reliance Foundation and UBS Optimus Foundation as Founding Partners, the British Asian Trust as the Programme Leader and Central Square Foundation as the Design and Technical Partner.

The Accelerator aimed to catalyse the supply of contextually relevant and pedagogically sound learning solutions, generate compelling evidence on their efficacy, work with governments to enhance the efficacy of EdTech adoption and create public goods to address systemic challenges in the ecosystem.

The Accelerator aimed to support **eight high-quality EdTech solutions** for two years through impact-focused grant funding, dedicated mentorship and capacity-building support to unlock the full potential of the EdTech solutions. The solutions were onboarded into three cohorts, each addressing key challenges in the Indian EdTech ecosystem. The cohorts focused on:

1. Scale – products looking to discover and unlock new pathways to scale - ThinkZone

2. Engagement – products seeking strategies to deepen engagement with the users - Chimple, Ei Mindspark, Pratham, Rocket Learning, Top Parent

3. Product Contextualisation – products developing pedagogically sound and contextually relevant solutions specifically for low-income India - Amira Learning and Sesame Workshop India (SWI)

On the demand side, the Accelerator focused on driving the adoption and institutionalisation of tech-based home learning for FLN within State Governments, while also exploring innovative pathways for EdTech integration through partnerships with retail channels, such as gig economy organisations and self-help groups (SHGs).

To tackle the challenge of limited existing evidence on 'what works' in EdTech and to allow for ongoing innovation and progress, the Accelerator's evidence generation agenda included

1. Learning Outcomes Evaluation - to assess the impact on student learning outcomes for ThinkZone, Top Parent and Ei Mindspark.

2. Impact of Acceleration Study - to capture the effectiveness of the strategies implemented within the Accelerator (published)

3. Insights on User Experience Study - a qualitative analysis that gathers feedback from end users on key aspects of the EdTech programme lifecycle, including acquisition, onboarding, engagement and retention.

These evaluations have been conducted under the supervision of the Principal Investigator, Prof. Tarun Jain (Reserve Bank of India Chair Professor of Economics at Indian Institute of Management, Ahmedabad) by experts from Sambodhi Research (qualitative study) and Educational Initiatives (quantitative study) to provide actionable insights to inform future interventions and improvements.

Between 2023-2025, the LiftEd EdTech Accelerator collaborated with eight leading EdTech organisations to advance the future of tech-enabled at-home learning for FLN in India. Collectively, these solutions have reached over **5.5 million users**, with more than **3 million** directly benefiting from features and innovations developed through the Accelerator's support.

This report provides insights from the **Learning Outcomes Evaluation Study** for **Ei Mindspark**.

2.3 Overview of EdTech Partner Model

Founded in 2001, Educational Initiatives (Ei) aims to ensure that every child learns with understanding. Ei collects large-scale student assessment data and utilizes it to power their technology-driven learning solutions. One of their EdTech products is Mindspark, an app launched in 2010. It is a personalized and adaptive learning software, which, through its gamified interface of questions, engages learners from low-resource settings and aims to improve their learning outcomes. Its content is aligned to the National Curriculum Framework. The app is available in and contextualised to nine Indian regional languages including Hindi, Kannada, Telugu, Marathi, Gujarati, Punjabi, Tamil, Odia, Urdu. Mindspark can be delivered in a variety of settings including in schools, in after-school centres, or at-home. Further, the web-app is platform-agnostic and can be deployed through computers, tablets, smartphones, or through a browser. While the app was originally deployed through schools, during the pandemic in 2020 Ei adapted the Mindspark model from a school based one to an at-home model. This study was conducted on Mindspark's at-home model.

The web-app features an item bank of over **45,000 test questions**, iterated over several years of design and field testing. This content can be uniformly and directly delivered to the child on parent's smartphones without requiring a parent or teacher to enable the usage. The content is adaptive with questions presented to students based on their learning levels. This adaptation is dynamic with every subsequent question completed, enabling "teaching at the right level". Additionally, it uses its large database of millions of observations of student's responses to questions to identify patterns of student errors and to classify the type of error and target differentiated remedial instruction accordingly. Using content with the above features, the app utilises an interactive user interface to engage children and boost their attention.

Mindspark follows a community-based engagement model. The web-app is anchored by the fieldworker and accessed by the children on their parent's smartphones to enable learning at-home and outside of school. The fieldworker, who typically belongs to a local NGO, approaches a local primary school and engages with the community in that school's vicinity. Contingent on the school's cooperation with them, they receive phone numbers and addresses of students from grade 1-3 in the area from the school's database. Following this, they introduce children and their parents to the application either in the school and/or in their homes (through door-to-door visits). If the school does not cooperate, the fieldworker directly reaches out through door-to-door visits to onboard children from Grades 1-3. With the consent of the parents, the children are onboarded on the platform. The fieldworkers facilitate and monitor the child's usage of the web-app. This is done through two mediums: a WhatsApp group, in which screenshots of completed activities is shared and regular field visits either to the house of the individual children or community spaces (such as outside a temple, a house in the neighbourhood) where they interact with small groups of the children who have been onboarded. They may or may not provide the children with devices (e.g. tablets, phones) during these visits. On a daily basis, the fieldworker assigns children activities based on their school grade.

The effectiveness of Mindspark has been tested through several studies. Findings of the studies have consistently shown greater learning outcomes for children using FLN Mindspark vis-à-vis children in comparison groups. A study by J-PAL[16] established that using Mindspark led to 2-4 times gain in learning outcomes among students between Grade 6 to 9 who were going to Mindspark centers in Delhi focused on serving low-income neighborhoods. However, findings show that the solution was not as effective for early grades. In response to this, their work under the aegis of the accelerator has attempted to incorporate skills-based learning and methodology, wherein the web-app focuses on identifying and developing skills through conceptual and associative learning. Additionally, through the accelerator, Mindspark is increasing nudges to drive 'use' behaviour by parents and children (through physical and frequent nudges).

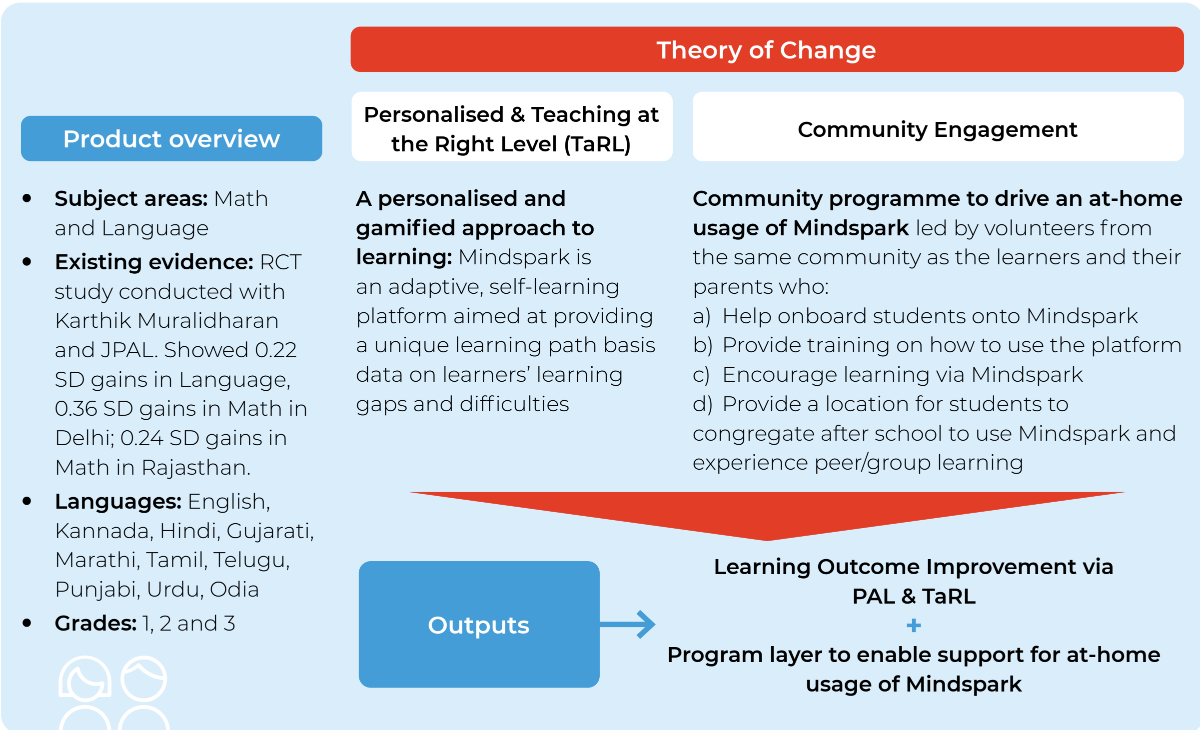

Figure 1: Ei Mindspark Model

Overall, the theory of change for Ei Mindspark integrates personalised, adaptive learning with community-based engagement to support both in-school and at-home use, with the aim of improving learning outcomes for early-grade students.

16 Karthik Muralidharan, Abhijeet Singh, Alejandro Ganimian (2018). Disrupting Education? Evidence on Technology-Aided Instruction in India. https://www.aeaweb.org/articles?id=10.1257/aer.20171112

3. Evaluation Design and Approach

This section outlines the study's research design, detailing how the quasi-experimental setup, sampling approach and assessment methods were structured to measure Ei Mindspark's impact on numeracy outcomes. It also summarises the timelines, data collection processes and safeguards used to ensure validity and reliability of findings.

3.1 Study Design

3.1.1 Overview

The study employed a quasi-experimental design (QED), with an Intervention group and a Comparison group identified to benchmark impact. This study seeks to answer the question:

> What impact does the use of the intervention programme have on student learning outcomes?

The Intervention group included students who received structured access to the Ei Mindspark programme through their teachers at schools and community centres. The Comparison group comprised similar students who did not receive the programme during the study period. This setup enabled a clear assessment of the effect of EdTech exposure by comparing progress between the two groups with comparable baseline characteristics. To estimate causal impact, the study used a difference-in-differences (DiD) approach, measuring changes in learning outcomes over time while controlling for time-invariant factors.

Students from both Grades 2 and 3 were included in the study, although sampling was not conducted at the grade level. Assessments were carried out in two rounds to measure learning outcomes at the start (baseline) and end (endline) of the evaluation period.

3.1.2 Timelines

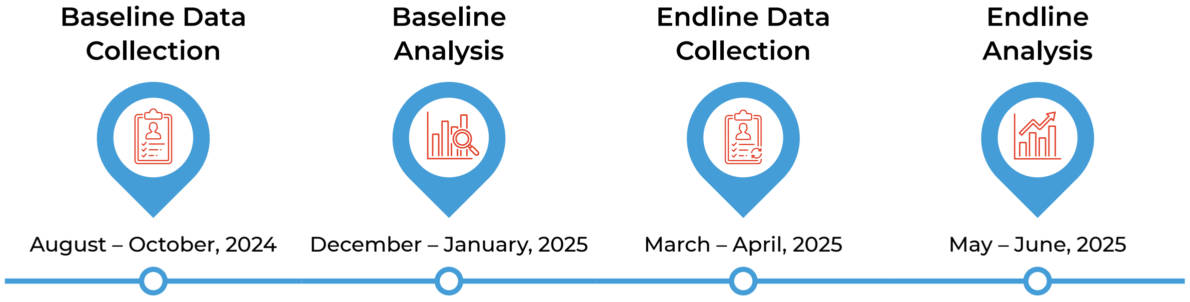

Figure 2: Study Timelines

3.2 Sampling Strategy

3.2.1 Sample Design

The evaluation was conducted in Ajmer, Bundi and Kota districts of Rajasthan wherein 12 blocks (7, 3 and 2 respectively) were identified as the intervention group and all other blocks utilised as the comparison group. Since the intervention group was pre-selected, the study's sample size was calculated based on the population of this group.

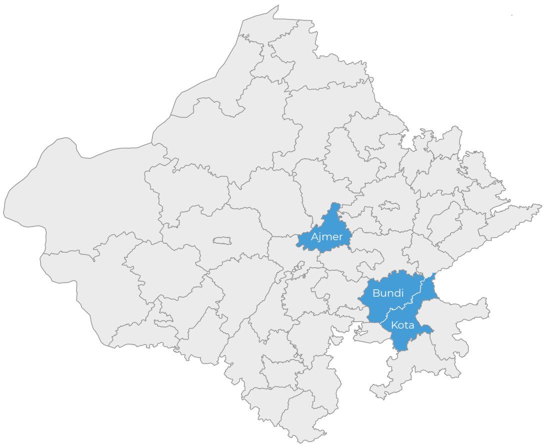

Figure 3: Geography covered

The following considerations were used for sample size calculation:

Core Statistical Inputs

95% Confidence Level
85% Power
MDES: 0.15
ICC: 0.10

Sample Size Parameters

Operational Adjustments

25% Absenteeism
25% Attrition*
30% Estimated ToT Effects*

Note: Factors marked with * apply only to the Intervention group.

Figure 4: Sample Size Parameters

Based on the above factors, the final sample size for both the Intervention and Comparison groups was determined as **1821 students from 93 Intervention schools and 1200 students from 60 comparison schools.** Details regarding the sampling process are presented in Annexure 6.1 to 6.3.

3.2.2 Selection of Intervention and Comparison Groups

Mindspark recruited students for its EdTech Accelerator intervention through two models:

- School-based approach: Relevant schools were identified where the team obtained permission to implement the programme and encourage students in the target grades to use the Mindspark FLN solution at home. Intervention group students were sampled from these schools.
- Community-based approach: Mindspark partnered with community organisations to onboard students from local neighbourhoods through door-to-door enrolment. These students, drawn from a range of nearby schools, were encouraged by either the community partner or the Mindspark team to use the FLN solution at home.

For the school-based model, comparison schools were selected from the same districts but outside intervention schools, using Coarsened Exact Matching (CEM) on socio-economic and demographic factors. While the initial plan was to select Comparison schools from different blocks to avoid spillover, the spread of intervention schools across 12 blocks in three districts made this infeasible. To mitigate risks of spillover within the same block, the evaluation sample size was powered at 5% above the original target. For the community-based model, the nearby government schools attended by onboarded students were treated as proxy Intervention schools for identifying matched comparison schools, again using the CEM process.

Given imbalances in Intervention and Comparison group sizes, many-to-one matching was applied to maximise overlap, allowing multiple Intervention schools to be matched with a single Comparison school while keeping operational costs manageable. Further details are presented in Annexure 6.2.

3.2.3 Final Sample Achieved

Since the study was longitudinal, tracking the performance of the same students over the evaluation period, students were excluded from the final dataset if they fell into one of the following categories at the end of data collection:

- Attempting only one subject in a given round due to lack of consent, operational challenges, or other constraints
- Absent or unsynced audio response recordings / assets
- Presence in the Baseline but absence in the Endline

The final sample is as follows:

Table 1: Details of Final Sample

Arm	Sample Size	Minimum Required	Final Sample Achieved
Intervention	1821	433	1085
Comparison	1200	433	980

Since the minimum required sample size was achieved, no changes were needed to the MDES of the study. A detailed summary of the sampling process is provided in Annexure 6.3.

3.3 Data Collection Methods and Tools

This section outlines the assessment tools, data collection processes and safeguards employed to ensure accuracy and reliability of findings.

3.3.1 Assessment Tool

A contextualised version of the Early Grade Mathematics Assessment (EGMA) was used for this study. The same assessment tool was administered to students in Grades 2 and 3 and remained unchanged across both the Baseline and Endline rounds. This was feasible given the EGMA-based FLN tool assesses largely common concepts and skills across Grades 2 and 3, with the exception of multiplication and division facts, which are typically introduced in Grade 3. The trade-off of excluding these Grade 3–specific skills is outweighed by the advantages of using a common tool, which enables the calculation of consolidated scores and effect sizes across both grades. This approach allowed the study to examine the impact of the EdTech solution across two grade levels while expanding coverage without increasing sample size or implementation costs. To ensure contextual relevance, the assessment tools were adapted, contextualised and translated into Hindi.

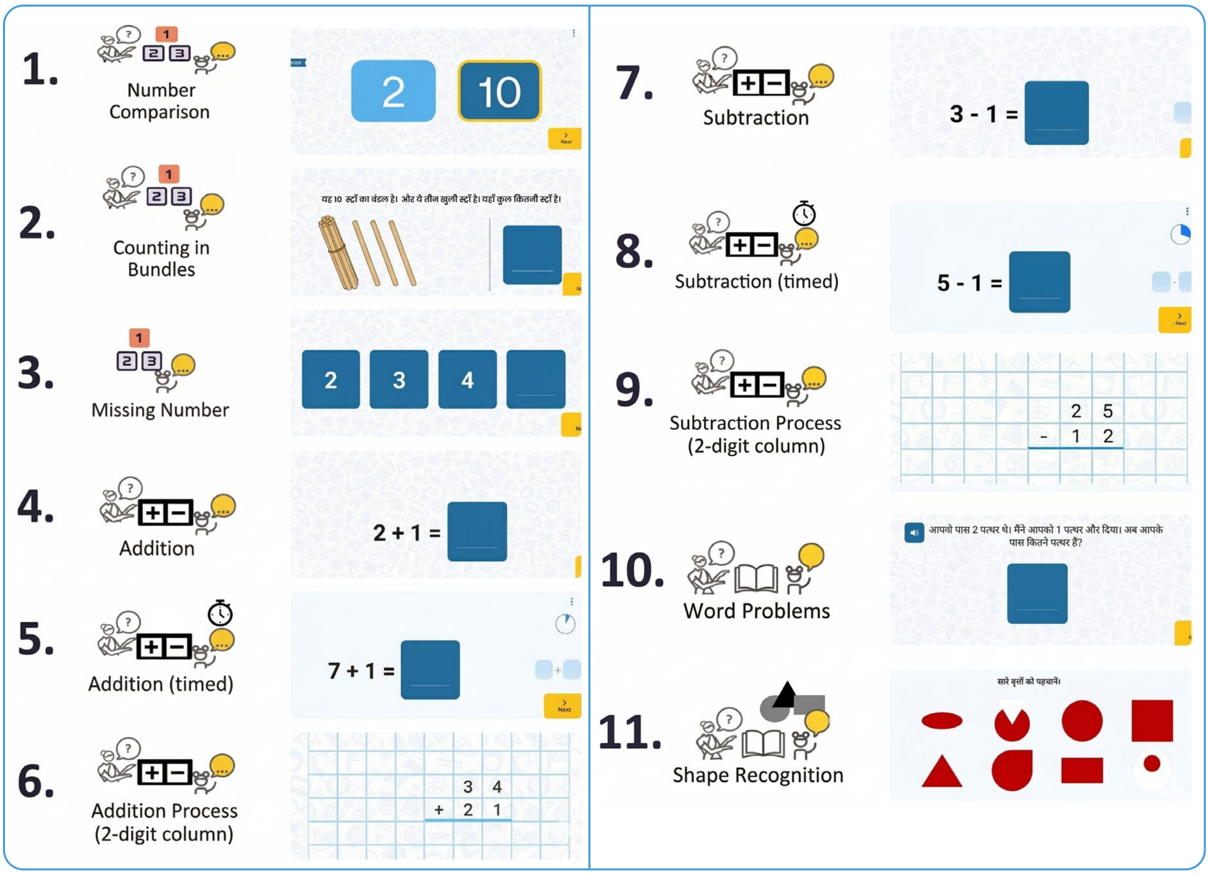

Figure 5: Tasks Assessed

The scaffolded assessments capture foundational skills, ensuring robust measurement of learning outcomes. Further details on the tool are provided in Annexure 6.4.

3.3.2 Data Collection Process

Data collection relied on a combination of in-school and after-school assessments. Wherever feasible, assessments were conducted within schools (with official permissions) rather than door-to-door, to minimise costs and logistical challenges. While each assessment was one-on-one in nature - with students responding independently on their own devices - each enumerator simultaneously supervised two children at a time to ensure smooth administration and adherence to protocols. Dedicated enumerator training sessions were conducted prior to each assessment round to standardise procedures and maintain data quality.

Assessments were administered using the Ei NEEV application, installed on tablets provided to students at the start of each session. The application included verbal instructions to guide students through the process. Data was captured as touch-based digital entries, which were auto-scored by the application.

3.3.3 Regional Considerations

The operations plan was designed to account for region-specific factors that could affect student availability and participation. Scheduling of assessments took into consideration public holidays, the local harvest season and the school academic calendar. These adjustments ensured that data collection was minimally disruptive, logistically feasible and allowed for smooth test administration across all sampled schools.

3.3.4 Safeguards for Data Quality

Data collected via the Ei NEEV application was stored in a central database and subsequently transferred to a secure evaluation portal. Tasks were processed using pre-validated scripts that had undergone Quality Analysis testing. For timed tasks, items not reached by students were automatically scored as zero to maintain consistency. Once both Baseline and Endline scoring was completed, further verification checks were conducted, including:

- Analysing scores across competencies[17]
- Checking progression patterns[18]
- Identifying anomalies such as outlier distributions[19]

3.4 Challenges and Implications

The study provides relevant evidence on the intervention's impact, though certain limitations mentioned below should be kept in mind when interpreting the findings.

3.4.1 Technical Challenges

- Infrastructure constraints in some schools limited the ability to seat students far apart, occasionally affecting the quietness of the assessment environment. However, verification of the audio files confirmed that the background noise had no impact on validity of the assessments.
- Additionally, during the Baseline, technical issues related to tablet–database synchronization also resulted in the loss of certain audio files. Although recovery efforts were made through system backups, a small portion could not be retrieved. By the Endline, these issues were largely resolved for the purposes of this programme, significantly reducing data loss. Given the large sample size achieved in the study, the minimum required sample size was comfortably exceeded, ensuring no risk to the validity of findings. A detailed breakdown of the final sample determination is included in Annexure 6.3.

3.4.2 Sample Attrition and Participation Gaps

Between the Baseline and Endline rounds, a proportion of students attritioned due to migration, absence from school, or unavailability during the data collection window. These cases were excluded from the final sample, reducing the overall sample size available for analysis.[20]

17 Aggregate and task-level scores were reviewed across competency domains to ensure internal consistency and expected variation in performance.
18 Baseline to Endline score changes were examined to verify plausible learning trajectories at the student and group levels.
19 Score distributions were screened to flag extreme values or irregular patterns that could indicate data or scoring issues.
20 Demographic Analysis of the students who dropped out of the study after the Baseline is presented in Annexure 6.5.

Lee Bounds analysis was conducted to assess the potential impact of student attrition between baseline and endline. As part of this analysis, students who were present at baseline but missing at endline were assigned a score of zero, thereby generating more conservative estimates of the intervention's impact and ensuring that the reported effects are robust to potential bias arising from differential dropout. Lee Bounds estimates[21] suggest that some of the positive DiD effects reduce in magnitude, lose statistical significance, or change direction once potential attrition bias is accounted for. This indicates that the main results should be interpreted with caution, as differential attrition may influence the estimated impacts.

3.4.3 Operational and Tool Related Challenges

In a few schools, assessments were completed before it was identified that the tool was incomplete on the application, resulting in students not attempting all intended sub-tasks. Consequently, revisits were conducted in these schools to ensure that students could attempt the full set of sub-tasks.

21 Lee bounds provide a conservative range of treatment effects by adjusting for potential bias from differential attrition, effectively assuming that all individuals who dropped out would have had the lowest possible outcome (e.g., scored zero).

4. Findings

This section presents the key findings from the evaluation, detailing the overall impact of Ei Mindspark on numeracy outcomes and how these effects vary by grade and gender. It also examines score movements to provide deeper insight into where the intervention drove progress and where challenges persist.

4.1 Overall Impact

The programme demonstrates **large, positive and statistically significant effects** (across all skills) on learning outcomes, with clear patterns that align with how numeracy skills develop. The **strongest gains appear in arithmetic operations and problem-solving domains**—particularly **Subtraction Facts** (timed, 0.81 SD), **Subtraction Process** (0.78 SD), **Addition Process** (0.69 SD), **Addition Facts** (timed, 0.69 SD) and **Shape Recognition** (0.71 SD). These effect sizes suggest that students not only strengthened their fact-retrieval fluency but were also able to apply these skills to multi-step operational tasks. For instance, the strong performance in timed Addition and Subtraction Facts complements the equally high gains in Addition and Subtraction Process, reflecting a coherent progression: rapid fact recall frees cognitive load, enabling students to carry out more complex procedures such as carrying and borrowing with greater accuracy and speed[22].

Similarly, gains in **Missing Number** (0.66 SD) and **Word Problems** (0.63 SD) indicate that improvements in computational fluency likely translated into better performance on tasks requiring reasoning and application of operations in context.

22 Cragg, L., Richardson, S., Hubber, P. J., Keeble, S., & Gilmore, C. (2017). *When is working memory important for arithmetic?* The impact of strategy and age. PLOS ONE, 12(12), e0188693. https://doi.org/10.1371/journal.pone.0188693. Cragg et al. (2017) explain that procedural arithmetic strategies require more working memory because children must hold interim results (and/or count steps) in mind while executing other steps, whereas fact retrieval is a single-step pull from long-term memory—so more fluent fact recall leaves more working-memory capacity available for multi-step procedures like column addition/subtraction with carrying/borrowing.

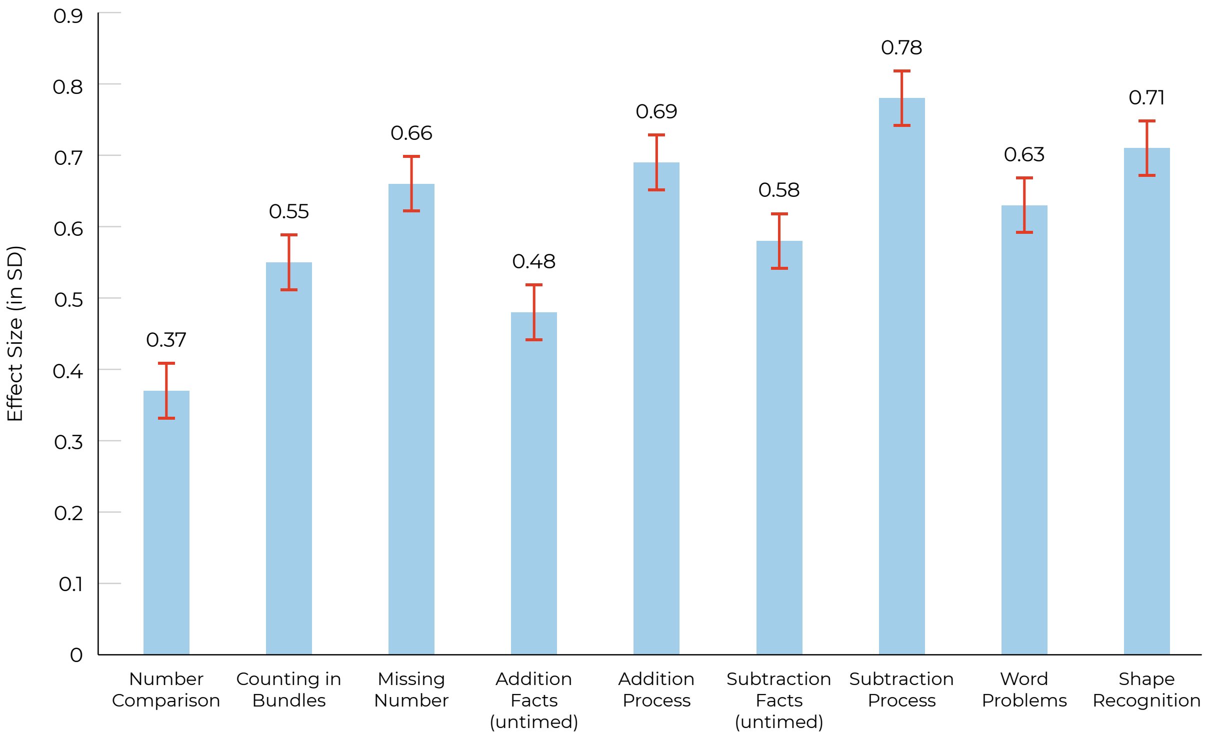

Figure 6: Task wise DiD Effect Sizes (in SD) – Untimed Tasks[23].

The red lines denote the standard errors.

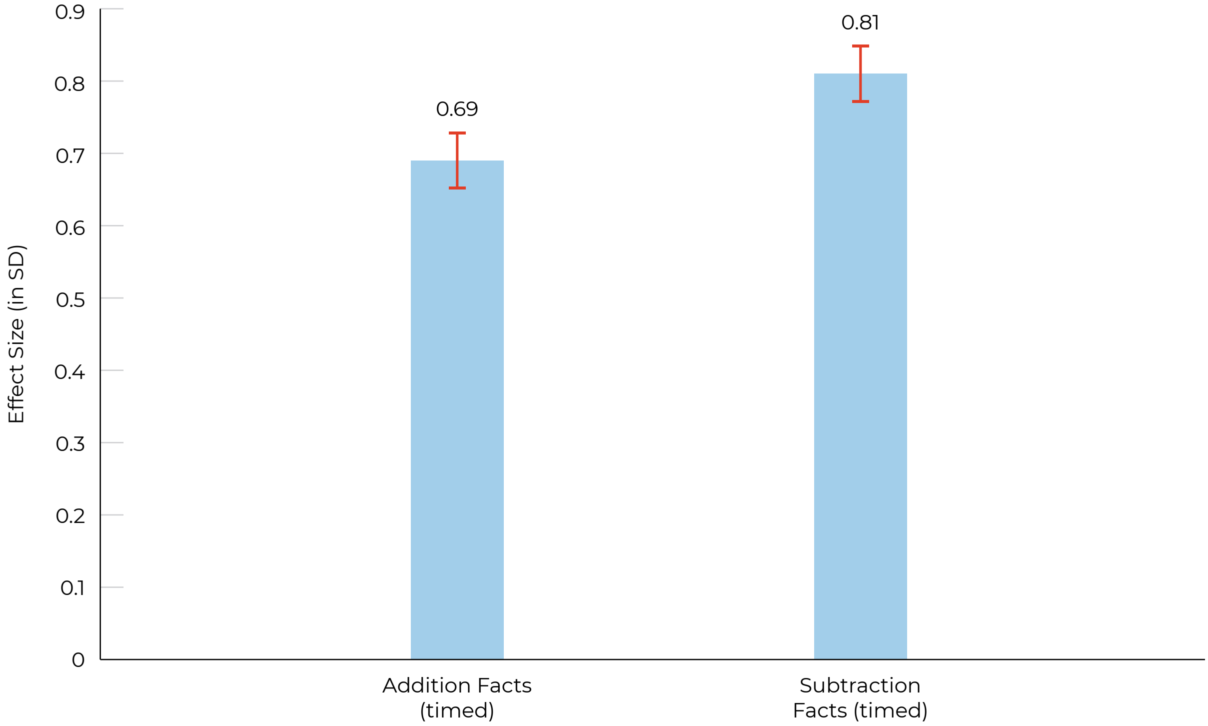

Figure 7: Task wise DiD Effect Sizes (in SD) – Timed Tasks.

The red lines denote the standard errors.

23 Detailed analysis has been shared in Annexure 6.6. Results of all sub-tasks are statistically significant i.e. $p < 0.05$.

Statistically significant impacts in **Subtraction Facts** (untimed, 0.58 SD), **Counting in Bundles** (0.55 SD), Addition Process (untimed, 0.58 SD) and **Addition Facts** (untimed, 0.48 SD) also align with the overall pattern. Counting in bundles shows place-value understanding, which helps children do multi-digit addition and subtraction - so the gains here support the improvements seen in conceptual operations. That is, as students better understand grouping and regrouping into tens, they become more adept at the carrying/borrowing steps embedded within addition and subtraction[24]. This coherence across indicators suggests that conceptual foundations and procedural fluency strengthened in parallel, reinforcing each other rather than developing in isolation.

Procedural competencies such as **Number Comparison** (0.37 SD) also showed smaller, but positive, improvements. This is reasonable within a skill-development trajectory: once students have already achieved basic number sense, incremental gains tend to be smaller compared to the larger jumps seen in procedural and problem-solving skills that benefit more directly from repeated practice, structured instruction and scaffolded tasks[25].

The pattern shows that the programme helps students build a strong foundation in number sense and pattern recognition. Its biggest impact is on improving arithmetic fluency, while also strengthening students' ability to solve problems, especially in subtraction, application-based tasks, and shape recognition. Overall, the programme effectively supports students gains in advanced numeracy skills.

4.2 Equivalent Years of Schooling (EYOS)

4.2.1 EYOS Computation Methodology

The Equivalent Years of Schooling (EYOS) metric is used to translate standardized numeracy learning gains into an interpretable measure expressed in years of schooling. For Rajasthan, EYOS is computed using data from a quasi-experimental design, where the control group represents **business-as-usual (BAU)**[26] instruction.

Learning gains are first calculated separately for intervention and control groups as the difference between endline and baseline mean scores. These gains are standardized using a pooled standard deviation to obtain effect sizes for each numeracy competency. EYOS is then computed using a **control-as-standard specification**, defined as the ratio of the intervention effect size to the control effect size. This ratio captures the extent to which learning progress among MindSpark students exceeds the BAU learning trajectory over the same period.

EYOS is estimated at the numeracy-competency level and aggregated across competencies using below specifications:

24 Jensen, Solveig, et al. "Place Value and Regrouping as Helpful Constructs to Diagnose Difficulties in Understanding the Place Value System." Journal für Mathematik-Didaktik, 2024, https://doi.org/10.1007/s13138-024-00234-8
25 Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. Child Development, 75(2), 428–444. https://doi.org/10.1111/j.1467-8624.2004.00684.x
26 Angrist, N., Bergman, P., Brewster, C., & Matsheng, M. (2020). Stemming learning loss during the pandemic: A rapid randomized trial of a low-tech intervention in Botswana. Journal of Human Resources, 56(S), S1–S45. https://doi.org/10.3368/jhr.58.S1.0620-12203R1 . Business-As-Usual (BAU) refers to the counterfactual learning trajectory that students would experience in the absence of the intervention, under prevailing instructional practices and system conditions. In impact evaluations, BAU is typically represented by outcomes observed in control or standard-implementation groups and serves as the benchmark against which incremental learning gains from an intervention are measured.

- **Weighted EYOS**, where competencies are aggregated using difference weights defined as the average of intervention and control sample sizes across baseline and endline. This specification gives greater weight to competencies with larger and more reliable samples and is treated as the primary estimate.

Treatment of Extreme Values and Winsorization

Because EYOS is a ratio-based measure, it is sensitive to extreme values arising from very small control-group effect sizes or noisy competency-level estimates. In the Rajasthan data, this issue is particularly relevant given heterogeneity across auto-scored numeracy competencies.

To ensure stability and interpretability of the EYOS estimates, the following rules are applied:

- **Control effect size floor:** Competency-level observations with control effect sizes below a minimum threshold are excluded from ratio calculations. This avoids artificial inflation of EYOS ratios driven by near-zero denominators that do not reflect meaningful BAU learning.
- **Upper-tail winsorization:** To prevent a small number of extreme competency-level ratios from disproportionately influencing the aggregate estimate, EYOS ratios are winsorized at an upper cap. Values above this cap are set equal to the cap rather than dropped, preserving the full set of competencies while limiting the influence of outliers.
- **Symmetric application:** These rules are applied consistently across weighted and unweighted specifications to ensure comparability.

The floor and cap thresholds are chosen based on inspection of the empirical distribution of effect sizes and are documented to ensure transparency and replicability. ***For Rajasthan, competency-level EYOS ratios were computed after excluding control effect sizes below 0.20 SD and applying an upper winsorization cap of 1.80× to limit instability from sparse or noisy competencies.***

Uncertainty around EYOS estimates is quantified using standard errors derived via the delta method, allowing for the construction of confidence intervals and hypothesis tests.

4.2.2 Findings

The EYOS analysis for MindSpark in Rajasthan indicates substantial learning gains in numeracy relative to BAU schooling.

Specification	Intervention: Control EYOS Ratio	EYOS Above BAU (Δ years)	Learning Acceleration (%)	Interpretation
Weighted EYOS	1.60×	+0.6 years	+60%	Preferred estimate; reflects learning gains aggregated with sample-size-based difference weights

In the table above we see that the control-as-standard EYOS ratio is estimated at **1.60**, implying that students in MindSpark intervention schools experienced learning progress equivalent to **60% faster growth** than their peers in control schools over the evaluation period. In schooling-equivalent terms, this corresponds to approximately **0.6–0.7 additional years of learning** beyond the BAU trajectory. This estimate is statistically distinguishable from BAU at the **95% confidence level**, indicating both meaningful magnitude and reasonable precision.

The EYOS results for Rajasthan indicate sustained and meaningful numeracy learning gains attributable to MindSpark, robust to alternative aggregation choices and clearly exceeding the learning trajectory observed under BAU instruction.

4.3 Impact by Grade

This section examines performance across grades to understand how the intervention affected students at different stages of learning and to highlight grade-specific trends in numeracy skills.

The grade-wise analysis of numeracy highlights strong improvements in both grades, with Grade 2 demonstrating larger and more consistent gains than Grade 3. This suggests that the intervention may align particularly well with the learning needs and cognitive readiness of early-grade students.

In untimed tasks, Grade 2 recorded large and **significant effects across almost all domains**, particularly Subtraction Process (0.93 SD), Addition Process (0.81 SD), Shape Recognition (0.72 SD) and Word Problems (0.70 SD). These effects appear mutually reinforcing: **strong gains in fact-based tasks**—such as Subtraction Facts (0.64 SD) and Addition Facts (0.57 SD) - **correspond with similarly large gains in the multi-step Process tasks**, which draw directly on accurate fact recall. Counting in Bundles (0.65 SD) shows gains as well, aligning with the improvements in Addition and Subtraction Process. Similarly, large improvements were also seen in Missing Number (0.82 SD), Word Problems (0.70 SD) and even procedural tasks like Number Comparison (0.47 SD), indicating **broad strengthening across both simple and complex numeracy skills**.

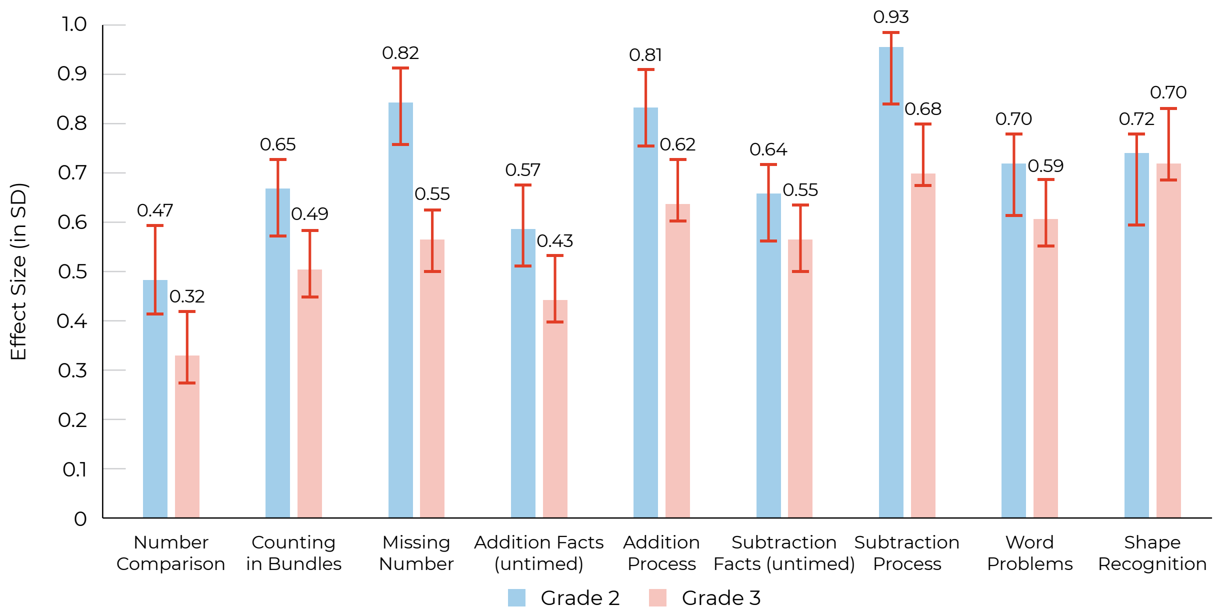

Figure 8: Grade-wise DiD Effect Sizes (in SD)– Untimed tasks[27].

The red lines denote the standard errors.

27 Detailed analysis has been shared in Annexure 6.7. Results of all sub-tasks are statistically significant i.e. $p < 0.05$.

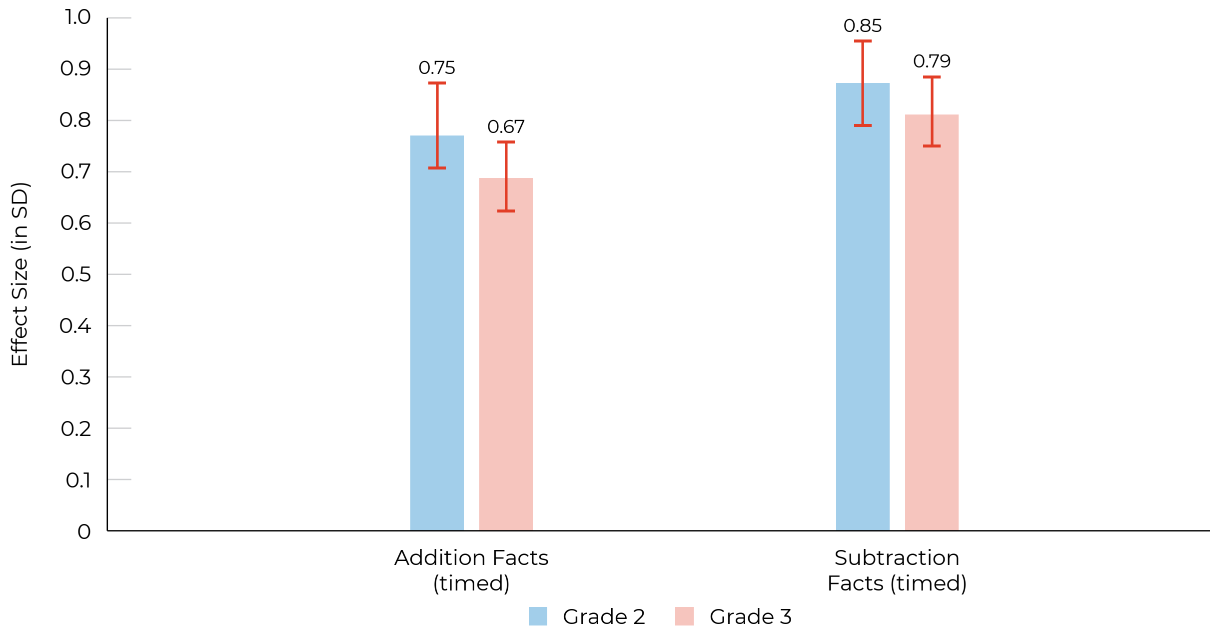

Figure 9: Grade-wise DiD Effect Sizes (in SD) – Timed tasks.

The red lines denote the standard errors.

Grade 3 also demonstrated significant positive effects, particularly in conceptual skills[28]. The largest effects appeared in Subtraction Process (0.68 SD), Shape Recognition (0.70 SD), Addition Process (0.62 SD) and Missing Number (0.55 SD), with progress in Word Problems (0.59 SD) as well. **Procedural skills** like Number Comparison (0.32 SD) also improved, though at **lower levels than Grade 2**. Counting in Bundles (0.49 SD) also improved, consistent with the gains in conceptual tasks where handling tens and ones is essential, though the magnitude is smaller than in Grade 2, mirroring the wider trend of comparatively lower gains in the higher grade[29].

Timed tasks reinforce this same cluster of patterns. Both grades show strong effects in Subtraction Facts and Addition Facts, which correspond with the large untimed gains and procedural tasks improvements - but again with Grade 3 at a smaller magnitude, reflecting the same ordering seen across other skills.

Overall, Grade 2 demonstrated substantially stronger gains than Grade 3 across both untimed and timed numeracy tasks, with consistently high effects in foundational skills, core arithmetic processes, and applied reasoning. The breadth and depth of improvement indicate that younger students responded more strongly to the intervention's design, benefiting from accelerated growth in both basic fluency and higher-order problem-solving.[30]

28 Across the conceptual skills (Addition Process, Subtraction Process, and Word Problems), Grade 3 shows slightly higher absolute performance than Grade 2 at both baseline and endline in the intervention group (e.g., endline: Addition Process 54% vs 49%, Subtraction Process 42% vs 39%, Word Problems 48% vs 45%; Tables 13–14). However, the programme's impacts on these skills are somewhat larger in Grade 2 (0.70–0.93 SD) than in Grade 3 (0.59–0.68 SD), partly because Grade 3 control groups also improved more on these outcomes (e.g., Addition Process control gain +20% in Grade 3 v. +16 % in Grade 2).

29 To examine whether the comparatively smaller gains observed in Grade 3 may be attributable to ceiling effects, baseline performance was reviewed for the relevant outcomes (Annexure 6.7). Baseline Intervention group averages indicate substantial room for improvement, and therefore limited risk of ceiling at the outset: Counting in Bundles is 9%, Addition Process is 15%, Subtraction Process is 7%, and Word Problems 15%. Even for Number Comparison, baseline performance remains below ceiling (58% in Grade 3; 46% in Grade 2). Endline averages for these conceptual outcomes also remain well below ceiling (e.g., Grade 3 Addition Process 54%, Subtraction Process 42%, Word Problems 48%), further suggesting that the Grade 3 pattern is unlikely to be driven by the assessment reaching its upper limit, and is more plausibly related to higher starting levels and concurrent improvement in the control group.

30 T-tests were run between Grade 2 and Grade 3 students

4.4 Impact by Gender

This section examines performance differences between boys and girls to identify potential gender gaps in learning outcomes and assess whether the programme benefits were equitably distributed[31].

The intervention generated positive impacts for both boys and girls, though the pattern of gains differed by gender. Boys' improvements were more pronounced in procedural tasks, while gender gaps narrowed considerably in pure arithmetic tasks. Girls registered consistent gains across all domains though the magnitude of these gains was smaller for girls in several tasks. Overall, the findings suggest that the intervention translated into measurable learning gains for students of both genders, albeit through different performance profiles.

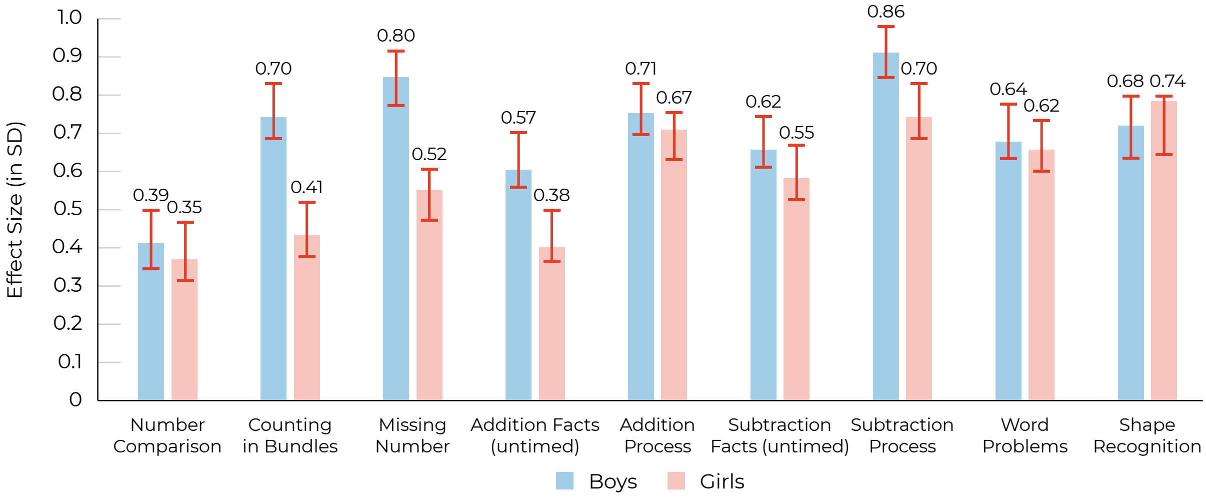

Figure 10: Gender-wise DiD Effect Sizes (in SD) – Untimed tasks[32]

The red lines denote the standard errors.

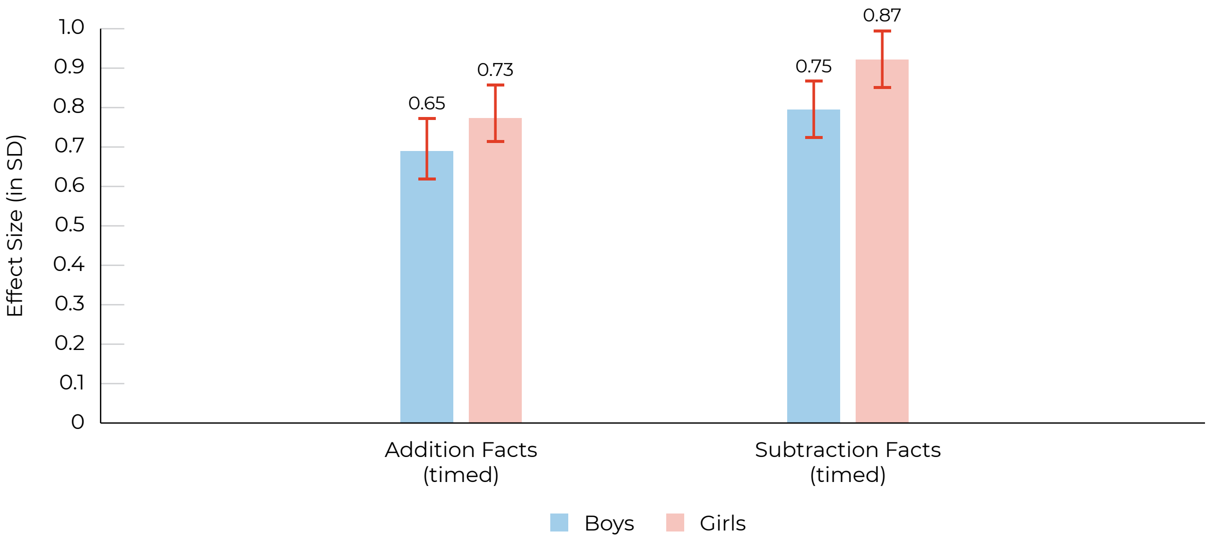

Figure 11: Gender-wise DiD Effect Sizes (in SD) – Timed tasks

The red lines denote the standard errors.

31 Boys and girls were comparable in their numeracy skills at study onset, with 9 out of 11 measures demonstrating gender equivalence ($p>0.05$). Only 2 numeracy subtasks showed significant baseline differences: Number Comparison – Untimed and Counting in Bundles ($p = 0.01$ each).

32 Detailed analysis has been shared in Annexure 6.8. Results of all sub-tasks are statistically significant i.e. $p < 0.05$.

Across most untimed tasks, boys showed larger DiD effect sizes than girls, with the exception of **Shape Recognition**, where girls demonstrated higher gains (0.74 SD vs. 0.68 SD). Gender differences were lower for **Subtraction Facts (Untimed)** (0.62 SD for boys vs. 0.55 SD for girls), but **larger** for **Addition Facts (Untimed)** (0.57 SD vs. 0.38 SD). For conceptual tasks, boys also show higher gains—**Subtraction Process** (0.86 SD vs. 0.70 SD) and **Addition Process** (0.71 SD vs. 0.67 SD)—suggesting comparatively stronger improvement in applying multi-step procedures. Boys' substantially higher gains in **Counting in Bundles** (0.70 SD vs. 0.41 SD) are consistent with this pattern, as place-value bundling is closely related to accuracy in multi-digit procedural work.

However, in timed tasks, **Girls actually outperform boys on timed fact-fluency tasks** - Subtraction Facts (0.87 SD vs. 0.75 SD) and Addition Facts (0.73 SD vs. 0.65 SD)- showing that speed and automaticity improved strongly for both, but more so for girls.

Overall, the patterns indicate a gender difference: boys demonstrated broader and larger gains across untimed tasks, particularly in higher-order problem-solving and procedural tasks, whereas girls' strongest outcomes were in timed arithmetic tasks. This suggests that while the programme effectively strengthened numeracy skills for all students, boys were better able to extend these gains into problem-solving, while girls excelled more in fluency-oriented tasks.

4.5 Score Distribution and Movement Analysis

This section presents the distribution of students across performance bands over assessment rounds, along with observed progressions and declines in student scores. Student performance at the sub-task level has been classified into five levels (L0–L4), defined separately for accuracy-based and fluency-based tasks, as outlined below.

Table 2: Categorisation for Accuracy Tasks

Category	Accuracy Tasks (%)	Fluency Tasks (CPM - Correct Per Minute)
L0	0%	0
L1	0 to 25%	0 to 0.25 times Baseline Average (BL Avg.)
L2	25 to 50%	0.25 to 0.5 times BL Avg.
L3	50 to 75%	0.5 to BL Avg.
L4	75 to 100%	> BL Avg.
Note: All category lower limits are inclusive, except for L1; for example, 25–50% includes 25% but excludes 50%.		

These categories enable tracking of the magnitude and direction of change in student performance across the evaluation timeline.

The tables that follow present the percentage distribution of students across performance bands at Baseline and Endline for each assessed sub-task. Percentages reflect the proportion of students in each category at the respective assessment round, such that L0 + L1 + ... + L4 = 100% for Intervention and Comparison groups[33]:

33 Detailed counts of students in each category, disaggregated by Intervention and Comparison groups, are provided in Annexure 6.9.

Table 3: Baseline Performance Category - Student Distribution

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Number Comparison (untimed)	8%	3%	11%	12%	18%	8%	2%	9%	12%	17%
Counting in Bundles (untimed)	44%	0%	4%	3%	2%	37%	0%	5%	2%	3%
Missing Number (untimed)	25%	6%	13%	6%	3%	20%	6%	11%	6%	4%
Addition facts (untimed)	24%	0%	7%	6%	15%	21%	0%	6%	6%	13%
Addition facts (timed)	17%	0%	6%	8%	22%	14%	0%	5%	7%	21%
Addition Process (untimed)	34%	6%	6%	6%	0%	32%	6%	4%	5%	1%
Subtraction facts (untimed)	27%	0%	13%	5%	8%	24%	0%	12%	5%	7%
Subtraction facts (timed)	25%	0%	6%	5%	16%	22%	0%	5%	4%	17%
Subtraction Process (untimed)	42%	5%	3%	3%	0%	39%	4%	2%	2%	0%
Word problem (untimed)	27%	16%	6%	3%	1%	26%	13%	5%	3%	1%
Shape Recognition (untimed)	49%	0%	3%	0%	0%	42%	0%	5%	0%	0%

Across almost all number and operations sub-tasks, the Intervention group shows a strong shift out of L0 into higher bands by endline, with especially large gains in automaticity tasks (i.e. timed addition/subtraction facts). Comparison schools improve too, but the upward shift is smaller and more mixed, particularly on process/word-problem items, where low-band concentrations remain higher than in Intervention.

Table 4: Endline Performance Category - Student Distribution

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Number Comparison (untimed)	6%	0%	3%	7%	36%	8%	1%	4%	9%	24%
Counting in Bundles (untimed)	14%	0%	8%	7%	24%	22%	0%	7%	5%	14%
Missing Number (untimed)	4%	3%	12%	12%	22%	10%	5%	13%	9%	10%
Addition facts (untimed)	5%	0%	4%	6%	37%	11%	0%	5%	7%	23%
Addition facts (timed)	3%	0%	2%	4%	44%	8%	0%	2%	6%	31%
Addition Process (untimed)	7%	5%	9%	16%	16%	17%	7%	8%	12%	4%
Subtraction facts (untimed)	5%	0%	7%	8%	32%	13%	0%	10%	7%	17%

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Subtraction facts (timed)	6%	0%	2%	2%	43%	13%	0%	3%	3%	28%
Subtraction Process (untimed)	13%	7%	7%	16%	10%	23%	9%	6%	7%	2%
Word problem (untimed)	8%	10%	7%	15%	13%	17%	10%	7%	10%	4%
Shape Recognition (untimed)	35%	0%	13%	4%	0%	41%	0%	6%	0%	0%

In addition to examining the distribution of students across performance bands at Baseline and Endline, this study also tracked changes in individual student performance over time. As a longitudinal study involving the same students, this allows for an analysis of broad trends in movement across performance bands between the two rounds.

The score movement analysis highlights clear contrasts between the Intervention and Comparison groups. The **Intervention group showed stronger upward mobility, despite having a lower share of students in the L0 category at baseline**, with the most pronounced shifts seen in procedural arithmetic skills. For instance, in Addition Facts (timed), 91% of L0 performers in the Intervention group advanced to L3 or L4, compared to 71% in the Comparison group. Similarly, in Subtraction Facts (timed), 87% in the Intervention group moved up versus 69% in the Comparison group. By contrast, far more students in the Comparison group remained at the lowest level: 23% and 30% in L0 for these tasks, compared to only 6% and 13% in the Intervention group.

Table 5: Score Movement Analysis[34]

Upward movewwment >10% Downward movement >10%

Addition Facts (Timed)					
Baseline Performance Category	Endline Performance Category				
	L0	L1	L2	L3	L4
Intervention					
L0	6%	0%	3%	8%	83%
L1	-	-	-	-	-
L2	8%	0%	5%	14%	73%
L3	6%	0%	5%	6%	83%
L4	5%	0%	2%	5%	88%
Comparison					
L0	23%	0%	6%	12%	59%
L1	-	-	-	-	-
L2	13%	1%	7%	15%	65%
L3	21%	0%	4%	11%	64%
L4	12%	0%	4%	11%	73%

Word Problems (Untimed)					
Baseline Performance Category	Endline Performance Category				
	L0	L1	L2	L3	L4
Intervention					
L0	16%	20%	13%	30%	22%
L1	15%	19%	14%	26%	25%
L2	12%	14%	17%	28%	29%
L3	9%	15%	9%	34%	32%
L4	5%	24%	5%	24%	43%
Comparison					
L0	40%	19%	13%	20%	8%
L1	35%	24%	15%	17%	10%
L2	24%	19%	22%	28%	6%
L3	32%	22%	8%	28%	10%
L4	10%	20%	30%	25%	15%

34 Detailed analysis has been shared in Annexure 6.9.3 and 6.9.4.

Mid-level performers also showed stronger upward mobility under the Intervention. In the Missing Number task, for example, 70% of students in the Intervention group advanced from L2 to L3 or L4, compared to only 42% in the Comparison group.

Conceptual skills reflected a more mixed picture. In Word Problems, performance was sustained among many Intervention students, with 43% of those in L4 and 32% of those in L3 maintaining their level. In contrast, stagnation was more common in the Comparison group, where 40% of L0 and 24% of L1 students remained at the same level by Endline. Some tasks even showed declines, most notably in Shape Recognition, where 100% of L3 performers in the Comparison group and 20% in the Intervention group fell back to L0 at Endline.

Overall, the score-movement patterns suggest that the Intervention supported stronger progression for students, with more learners moving out of the lowest performance levels, steady gains among mid-level performers, and generally stable performance in several conceptual skills. While a few tasks showed mixed or declining trends, the broader pattern indicates a positive shift in foundational and procedural numeracy. This is relevant because improvements in movement across performance levels can help narrow early learning gaps that often widen over time. By enabling more students to reach higher proficiency bands, the programme may be contributing to stronger numeracy foundations and better preparedness for subsequent mathematical learning.

5. Discussion and Implications

The evaluation of Ei Mindspark in Rajasthan demonstrates that the programme generated substantial improvements in foundational numeracy outcomes within the intervention window. Overall, the findings suggest that Mindspark supports both remediation and the development of deeper learning, with differentiated impacts across grades, gender and baseline performance levels.

- **The EYOS estimate of approximately 1.6 years indicates that the intervention enabled students to achieve learning gains equivalent to more than one and a half years of Business-As-Usual instruction over the study period.** This magnitude reflects substantial learning acceleration rather than incremental improvement, suggesting that the programme meaningfully compresses the time required for students to reach expected competency levels. Such gains are particularly consequential in low-learning contexts, where delayed foundational skills compound over time. The EYOS result underscores the potential of well-designed, high-intensity EdTech models to not only raise learning levels but also close accumulated learning gaps within a relatively short timeframe, strengthening the case for scale-up where system capacity and implementation fidelity can be sustained.
- **Strong gains across tasks, particularly conceptual skills**
  The programme produced its largest effects in conceptual competencies such as arithmetic fluency and non-rote problem-solving. This suggests that the adaptive design effectively supports students in progressing beyond memorisation towards conceptual understanding and application, positioning Mindspark as a tool for advancing deeper competencies at scale.
- **Larger improvements for Grade 2**
  Grade 2 students experienced wider and stronger effects than Grade 3 students, with particularly large gains in skills such as Subtraction Process and Word Problems. These results highlight the value of earlier exposure, indicating that integration at lower grades may yield stronger and more sustained learning trajectories.
- **Gender-differentiated learning patterns**
  Distinct gender-wise trends emerged, with girls showing stronger improvements in fluency-oriented tasks and boys exhibiting larger gains in untimed, accuracy oriented tasks. While both groups benefitted meaningfully, the divergence in skill-specific gains points to the potential value of differentiated instructional supports to balance fluency and reasoning outcomes.
- **Strong upward mobility among low performers**
  Score movement analysis reveals substantial upward progression among the lowest-performing students, underscoring the programme's levelling effect in rapidly strengthening foundational and applied skills. In contrast, progress among higher-performing students was more mixed, suggesting the need for differentiated pathways to sustain growth at the top end of the distribution.

6. Technical Annexures

6.1 Details of Sampling Design

- For the evaluation, using 95% confidence[35], 85% power[36] and 0.1225 variance[37], the required sample size is 355 students each in the Intervention and Comparison groups.
- Minimal Detectable Effect Size (MDES)[38]: The study aims to be able to detect a 0.15 SD difference between the baseline and endline rounds. The formula used to calculate MDES is:

$$MDE = (t_K + t_\alpha) \cdot \sqrt{\frac{1}{P(1-P)} \cdot \frac{\sigma^2}{N}} \cdot \sqrt{1 + (m-1) \cdot ICC}$$

MDE = Minimal detectable effect
t_K and t_α = Critical values from Student's t for power K and significance level α
σ^2 = Variance; N = Sample size; P = Proportion in treatment; m = Cluster size;
ICC = Intraclass correlation

- Since sampling by school introduces intra-class correlation (ICC[39]), a design effect (DEFF) adjustment was applied. Assuming ICC = 0.1 and an average cluster size of 20 students, DEFF was calculated as:

$$DEFF = 1 + 0.1 \times (20 - 1) = 2.9$$

- This increased the intervention and comparison sample size to 433 each.
- To account for absenteeism[40] (25% buffer), the sample increased to 577 per group. Factoring in 50% retention for a longitudinal study[41], the required sample became 770 per group.
- Additionally, to estimate the Treatment-on-Treated (ToT) effect, it was assumed 30% of intervention students would meet the ideal usage threshold[42]. This required sampling 2564 intervention students. Since the total intervention population is 1821, a full census was conducted for the intervention group.

6.2 Selection of Comparison Group

The comparison group for this quasi-experimental study would consist of schools that have not been given the treatment but are as similar as possible to the schools in the Intervention group in terms of baseline (pre-treatment) characteristics. The objective

35 The degree of certainty that the results reflect the true effect in the population rather than chance.
36 The probability of correctly detecting an effect if it truly exists.
37 The amount of natural variation in the outcome across individuals rather than uniform responses.
38 The smallest difference between groups that the study is designed to detect.
39 A measure of how similar students are within the same class or school, which affects the extent of independent information each student contributes. For this study, the intra-class correlation (ICC) was assumed to be 0.1, resulting in a Design Effect of 2.9.
40 The expected loss of sample due to students being absent during assessment, estimated at 25% for this study.
41 As the study was longitudinal and conducted over several months, some students present at baseline were unavailable at endline, hence an attrition buffer was included.
42 The estimated impact only on those who actually received the intervention as expected, estimated at 40% for this study.

of the comparison group is to capture what would have been the student learning outcomes in schools in the demonstration districts if the treatment was not given to these schools (i.e., the counterfactual). Hence, it is critical to identify the factors that are most likely to affect the outcome variable and then match the Intervention and Comparison observation units on these factors to create a valid counterfactual. While there are different techniques for creating a valid comparison group such as Propensity Score Matching, Regression Discontinuity Design and Mahalanobis Distance Matching, **Coarsened Exact Matching (CEM)** was used in this study for the following reasons:

1. A Monotonic Imbalance Bounding (MIB) matching method, CEM aims to balance the Intervention and Comparison groups ex ante
2. CEM also bounds through ex ante user choice both the average treatment effect estimation error and the degree of model dependence
3. CEM has been shown to produce good covariate balance between exposure groups and, thus, to reduce the impact of confounding in observational causal inference

6.2.1 Coarsened Exact Matching (CEM)[43]

The basic idea of CEM is to coarsen each variable by recoding it, so that substantially indistinguishable values are grouped and assigned the same numerical value. An exact matching algorithm is then applied to the coarsened data to determine the matches and to prune unmatched units.

In simple terms, the CEM process does not try to match each student in the Intervention group with a comparable student in the comparison group. Instead, it groups the characteristics of all the students in a particular school in the intervention group and then tries to find a school in the comparison group whose grouped characteristics exactly match the grouped characteristics of the original school. For example, let us assume that there is a school in the intervention group with 5 students in Grade 2, whose age (in years) and gender (M / F) is as follows:

- 6, M
- 6, M
- 6, M
- 7, F
- 7, F

The CEM method will coarsen this data into groups or bins and represent the school as one with less than 10 students in Grade 2, between 50% to 75% of whom are male and whose ages are either 6 or 7 years, with 50% to 75% of them being 6 years old. As a result, this school will be an exact match for the following school in the same district where the intervention is not being implemented and which has 9 students in Grade 2, whose age (in years) and gender (M / F) is as follows:

- 6, F
- 6, F
- 6, F

43 Iacus, S. M., King, G., & Porro, G. (2011). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1–24. https://doi.org/10.1093/pan/mpr013; Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. Journal of the American Statistical Association, 106(493), 345–361. https://doi.org/10.1198/jasa.2011.tm09599

- 6, F
- 6, M
- 6, M
- 7, M
- 7, M
- 7, M

Thus, the result of the CEM process in this case will be that each school in the sampling frame for the intervention group will be paired with a comparable school in the sampling frame for the comparison group and the sampling of the intervention and comparison groups will be done from these two sets of schools. To account for unforeseen challenges and situations on the ground during the data collection, a list of replacement schools will also be identified for the schools that make up the sampling frame for the comparison group.

6.2.2 Variables for the CEM Process:

Given below is the list of school-level variables that were used for the CEM process, to conduct the pairwise matching of schools, based on the data available on the UDISE website. These variables were chosen due to their potential relevance to student learning outcomes, informed by prior project experience.

Table 6: School-level variables

#	School-level variables
1	Total number of students in Grade 2
2	Total number of students in Grade 3
3	Pupil teacher ratio for the entire school
4	Location (urban vs. rural)
5	School (national) management type (e.g., dept. of education, local body, govt. aided, etc.)
6	Composite Index of Technological Infrastructure related variables (comprising of: count of functional laptops, desktops, tablets, scanners, printers, webcams, digi-boards, and access to the internet)[44]
7	School category (e.g., primary, primary with upper primary, etc.)

In scenarios where no exact match was present in the district for an intervention group school using all 7 variables mentioned above, some of these school-level variables were either coarsened further or dropped entirely in the subsequent rounds of the matching process, to ensure that a match was identified, while excluding the already matched intervention group schools. For example, in the absence of an exact match, the number of students in Grade 2 was coarsened into class intervals with a width of 20 students instead of 10 students for the next round of matching. However, the first 4 variables mentioned above were always retained in the CEM process.

44 All these variables are dichotomous in the UDISE dataset (meaning they have values of 1 for yes and 0 for no). The composite index was created by multiplying by taking the maximum value of the component variables, i.e., if any one of the variables had a value of 1, the composite variable would also have a value of 1.

6.2.3 Summary table of Matching Exercise

Based on the approach detailed above, a summary of the results of the matching exercise is given below:

Table 7: Summary of Matching Exercise

Matching Round	No. of unique intervention group schools	No. of unique comparison group schools	Details of matching variables
1	78	50	The count of students in Grades 2 & 3 and the pupil-teacher ratio, were coarsened into class intervals of width 10 from 1 to 100, with a separate class for 0 and then into intervals of width 50 from 101 to 300, with a separate class for any values above 300.
2	4	3	The school category variable was dropped.
3	1	1	The aggregated tech infra variable was dropped and the school category variable was re-introduced.
4	3	2	Both the aggregated tech infra and school category variables were dropped.
5	6	4	The width of the class intervals for the coarsened count of students in Grades 2 & 3 and the pupil-teacher ratio, was increased from 10 to 20 for values between 1 to 100. The aggregated tech infra and school category variables were also re-introduced.
6	1	0	The width of the class intervals for the coarsened count of students in Grades 2 & 3 and the pupil-teacher ratio, was increased from 20 to 50 for values between 1 to 100.
Total	**93**	**60**	

The maximum number of Intervention group schools that a single Comparison group school was matched with as a result of this process was 6. Additionally, given that the average class size of the Intervention group schools was 20 and the fact that all 60 matched schools in the Comparison group needed to be covered, the actual target sample to be assessed on the ground was 1,200 students for the Comparison group.

Note: The matching process used to identify the Comparison group schools had to be repeated multiple times, to identify replacements for those schools where assessments could not take place due to various issues - inaccessibility due to flooding, permission for the assessment being denied and a non-LiftEd Mindspark intervention being present. 47 Comparison group schools were replaced as a result of these re-sampling rounds, with an additional Comparison group school also being dropped because it was matched only with a Intervention group school where all the students targeted by the intervention were discovered to be Grade 4 students, in the updated student enrolment list shared by the Mindspark team.

6.3 Final Sample Achieved

The table below presents a detailed summary of the number of students assessed as part of the study, along with the number excluded from the analysis due to the following reasons:

- Attempting only one subject in a given round due to lack of consent, operational challenges, or other constraints
- Absent or unsynced audio response recordings / assets
- Presence in the Baseline but absence in the Endline

Table 8: Final Sample Achieved in Intervention and Comparison groups

Description	Intervention				Comparison			
	Baseline		Endline		Baseline		Endline	
	Literacy	Numeracy	Literacy	Numeracy	Literacy	Numeracy	Literacy	Numeracy
Total number of assessments administered	1221	1262	1290	1255	1057	1042	1072	1047
Students excluded due to operational constraints[45]	5	74	13	6	15	30	4	9
Students excluded due to technical issues[46]	79		139		21		53	
Students excluded as per study design criteria[47]	52		25		41		5	
Final Student Sample	1085				980			

45 Students who attempted only one of the two subjects within a given assessment round.
46 The student's assessment having absent or unsynced recordings/assets.
47 Students not present in both assessment rounds i.e. Baseline and Endline.

6.4 Modified EGMA Tool

The below table summarises the tool utilised in this evaluation.

Table 9: Details of modified Numeracy Tool and Skills Covered

Concept / Skill Assessed	Total Items/ Questions	Nature of Task	Metric	Objective	Mode of Evaluation
Number discrimination (Identify and read out the greater number among the given pairs of numbers)	10	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can compare and order numbers up to 1,000 accurately.	Automated
Counting in bundles (Give the total count of straws when given a few bundles of 10 straws each, along with some single straws. Also choose the correct count of bundles of 10 straws and single straws that add up to a given number.)	4	Untimed	Number of questions answered correctly	The objective of this task is to check if the child understands the concept of place value, i.e., if they understand how many tens and ones make a 2-digit number.	Automated
Number patterns (Identify a pattern in numbers and call out the missing number in each pattern)	8	Untimed	Number of missing numbers identified correctly	The objective of this task is to check if the child can recognize and complete patterns involving skip counting in 1s, 10s, 5s and 2s, for numbers up to 1,000.	Automated
Single-digit addition facts (Solve single-digit addition problems)	3	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can add single-digit numbers fluently.	Automated
	20	Timed: 1 minute			Automated
Multi-digit addition in vertical format (Solve multi-digit addition problems)	5	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can add multi-digit numbers accurately.	Automated
Subtraction facts within 18 (Solve subtraction problems involving numbers within 18 and with single-digit answers)	3	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can subtract numbers up to 18 and with single-digit answers, fluently.	Automated
	20	Timed: 1 minute			Automated

Concept / Skill Assessed	Total Items/ Questions	Nature of Task	Metric	Objective	Mode of Evaluation
Multi-digit subtraction in vertical format (Solve multi-digit subtraction problems)	5	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can subtract multi-digit numbers accurately.	Automated
Operations involving 0 (Solve addition and subtraction problems involving 0)	4	Untimed	Number of questions answered correctly	The objective of this task is to check if the child understands the concept of zero accurately. 1 question each in addition facts (untimed), addition process, subtraction facts (untimed) and subtraction process[48].	Automated
Word problems on single-digit number operations (Give an answer to narrated word problems)	6	Untimed	Number of questions answered correctly	The objective of this task is to check if the child can apply basic addition and subtraction in real-world scenarios.	Automated
Shape recognition (Identify a specific shape among a collection of shapes)	3	Untimed	Number of shapes recognized correctly	The objective of this task is to check if the child can recognize basic shapes like circles, rectangles and triangles. One question each on circles, rectangles and triangles. Each question involves showing the student multiple shapes (8, 10, and 11 respectively), of which only a few (2, 4 and 4 respectively) are correct examples of the shape the student has been asked to identify.	Automated

48 These items are already included in the count of the respective sub-tasks.

6.5 Demographic Analysis of Students Who Dropped Out after Baseline

Table 10: Demographic Analysis of Baseline Dropouts

Indicator		Intervention		Comparison	
		Students absent after Baseline	Overall Student Pool	Students absent after Baseline	Overall Student Pool
No. of Students		209	963	71	609
Student Age (avg.)		8.2 (8.2 - 8.3)*	8.4 (8.3 - 8.4)	8.3 (8.0 - 8.6)	8.3 (8.2 - 8.4)
Grade Split	Grade 2	42% (35% - 50%)	39% (35% - 42%)	51% (40% - 61%)	47% (36% - 58%)
	Grade 3	58% (51% - 65%)	61% (58% - 65%)	49% (35% - 63%)	53% (39% - 67%)
No. of Sisters (avg.)		1.4 (1.3 - 1.5)	1.4 (1.3 - 1.4)	1.3 (1.1 - 1.4)	1.3 (1.3 - 1.4)
No. of Brothers(avg.)		1.2 (1.1 - 1.2)	1.1 (1.1 - 1.1)	1.1 (1.0 - 1.2)	1.0 (0.9 - 1.0)
Socio-economic Context**	Low	7% (-4% - 18%)	5% (1% - 9%)	4% (-4% - 13%)	4% (-4% - 13%)
	High	93% (90% - 96%)	95% (93% - 96%)	96% (93% - 98%)	96% (94% - 98%)

* 95% Confidence Interval mentioned in brackets
* Binary flag created on the basis of appliances/transport available at home (i.e. feature phone v. smart phone; cycle v. motor cycle)

The analysis indicates that students absent after baseline in the Intervention group were largely comparable to the overall intervention pool across key demographic characteristics. The confidence intervals for average age show substantial overlap between absent students (8.2 years; CI: 8.2–8.3) and the overall pool (8.4 years; CI: 8.3–8.4), suggesting no meaningful age-based differences. Similarly, the CIs for the average number of sisters and brothers are nearly identical across groups, indicating comparable family composition.

Grade composition patterns also appear broadly aligned. The confidence intervals for the proportion of Class 2 students among those absent after baseline (42%; CI: 35%–50%) overlap with those of the overall intervention pool (39%; CI: 35%–42%), as do the intervals for Class 3 students. Socio-economic composition shows a similar pattern: although the share of high socio-economic households is marginally lower among absent students (93%; CI: 90%–96%) compared to the overall pool (95%; CI: 93%–96%), the overlapping CIs suggest close alignment.

In the Comparison group, demographic characteristics of students absent after baseline also closely mirror those of the overall student pool. The average age estimates are identical in point terms (8.3 years), with overlapping CIs, indicating no age-related differentiation. Grade-wise proportions for both Class 2 and Class 3 show wide and overlapping confidence intervals between absent students and the overall group, suggesting that observed differences in grade concentration fall within sampling variation. Family structure indicators and socio-economic status display almost complete overlap in confidence intervals, pointing to near-identical distributions.

Overall, the extensive overlap of confidence intervals across demographic indicators suggests that attrition after baseline is unlikely to be systematically associated with age, grade, family structure, or socio-economic background in either the Intervention or Comparison groups[49].

6.6 Overall Performance

Table 11: Detailed Results

Mindspark											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	76%	64%	54%	55%	23%	9%	36%	0.37	0.00
Counting in Bundles	Untimed	%	51%	35%	7%	11%	38%	17%	33%	0.55	0.00
Missing Number	Untimed	%	59%	39%	21%	23%	40%	21%	41%	0.66	0.00
Addition Facts	Untimed	%	74%	55%	34%	35%	5.53	2.08	4.98	0.48	0.00
Addition Facts	Timed	CPM	8.89	5.89	3.36	3.81	0.4	0.2	6.2	0.69	0.00
Addition Process	Untimed	%	52%	29%	13%	11%	42%	20%	38%	0.69	0.00
Subtraction Facts	Untimed	%	66%	45%	24%	25%	5.92	1.97	4.88	0.58	0.00
Subtraction Facts	Timed	CPM	8.0	4.6	2.1	2.6	0.3	0.1	6.3	0.81	0.00
Subtraction Process	Untimed	%	41%	19%	7%	6%	11%	1%	15%	0.78	0.00
Word Problems	Untimed	%	47%	28%	14%	13%	44%	24%	37%	0.63	0.00
Shape Recognition	Untimed	%	14%	5%	2%	4%	33%	15%	29%	0.71	0.00

49 It is important to note that demographic information was not mandatory during data collection and therefore, while representative, it may be incomplete.

6.7 Grade-wise Results

Table 12: Detailed Results - Grade 2

Mindspark - Grade 2											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	74%	60%	46%	49%	28%	11%	35%	0.47	0.00
Counting in Bundles	Untimed	%	50%	28%	5%	5%	45%	23%	35%	0.65	0.00
Missing Number	Untimed	%	58%	33%	16%	17%	42%	15%	32%	0.82	0.00
Addition Facts	Untimed	%	71%	46%	27%	25%	44%	21%	40%	0.57	0.00
Addition Facts	Timed	CPM	8.8	5.00	2.74	2.76	6.0	2.2	5.0	0.75	0.00
Addition Process	Untimed	%	49%	23%	9%	7%	40%	16%	30%	0.81	0.00
Subtraction Facts	Untimed	%	64%	38%	21%	18%	43%	20%	37%	0.64	0.00
Subtraction Facts	Timed	CPM	7.7	3.86	1.81	1.93	5.9	1.9	4.7	0.85	0.00
Subtraction Process	Untimed	%	39%	13%	6%	3%	34%	10%	26%	0.93	0.00
Word Problems	Untimed	%	45%	24%	12%	10%	34%	14%	28%	0.70	0.00
Shape Recognition	Untimed	%	13%	5%	2%	4%	11%	1%	15%	0.72	0.00

Table 13: Detailed Results - Grade 3

Mindspark - Grade 3											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	78%	68%	58%	60%	19%	8%	36%	0.32	0.00
Counting in Bundles	Untimed	%	52%	41%	9%	17%	44%	25%	38%	0.49	0.00
Missing Number	Untimed	%	59%	45%	23%	27%	36%	18%	33%	0.55	0.00
Addition Facts	Untimed	%	76%	63%	38%	43%	37%	20%	40%	0.43	0.00
Addition Facts	Timed	CPM	8.97	6.61	3.75	4.66	5.22	1.95	4.90	0.67	0.00

Mindspark - Grade 3											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value
			I	C	I	C	I	C			
Addition Process	Untimed	%	54%	34%	15%	15%	39%	20%	31%	0.62	0.00
Subtraction Facts	Untimed	%	67%	50%	26%	30%	41%	20%	38%	0.55	0.00
Subtraction Facts	Timed	CPM	8.17	5.15	2.25	3.14	5.92	2.02	4.96	0.79	0.00
Subtraction Process	Untimed	%	42%	23%	7%	8%	34%	15%	28%	0.68	0.00
Word Problems	Untimed	%	48%	31%	15%	16%	33%	15%	30%	0.59	0.00
Shape Recognition	Untimed	%	14%	5%	2%	4%	11%	1%	15%	0.70	0.00

6.8 Gender-wise Results

Table 14: Detailed Results - Boys

Mindspark - Boys											
Task	Task Type	Unit	Endline - Boys		Baseline - Boys		Delta - Boys		Pooled SD	Effect Size	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	78%	65%	56%	57%	22%	8%	36%	0.39	0.00
Counting in Bundles	Untimed	%	54%	35%	7%	14%	47%	20%	37%	0.70	0.00
Missing Number	Untimed	%	60%	40%	19%	25%	41%	14%	33%	0.80	0.00
Addition Facts	Untimed	%	74%	55%	33%	38%	41%	18%	41%	0.57	0.00
Addition Facts	Timed	CPM	8.94	6.12	3.41	3.83	5.53	2.29	4.98	0.65	0.00
Addition Process	Untimed	%	53%	30%	13%	12%	40%	18%	31%	0.71	0.00
Subtraction Facts	Untimed	%	68%	44%	24%	24%	43%	20%	38%	0.62	0.00
Subtraction Facts	Timed	CPM	7.90	4.77	2.07	2.50	5.83	2.26	4.78	0.75	0.00
Subtraction Process	Untimed	%	41%	18%	6%	6%	35%	12%	27%	0.86	0.00
Word Problems	Untimed	%	47%	28%	14%	14%	33%	15%	29%	0.64	0.00
Shape Recognition	Untimed	%	14%	5%	3%	4%	11%	1%	15%	0.68	0.00

Table 15: Detailed Results - Girls

Mindspark - Girls											
Task	Task Type	Unit	Endline - Girls		Baseline - Girls		Delta - Girls		Pooled SD	Effect Size	p-value
			I	C	I	C	I	C			
Number Comparison	Untimed	%	75%	63%	52%	53%	23%	10%	36%	0.35	0.00
Counting in Bundles	Untimed	%	49%	36%	7%	9%	42%	27%	36%	0.41	0.00
Missing Number	Untimed	%	57%	39%	22%	20%	36%	19%	32%	0.52	0.00
Addition Facts	Untimed	%	74%	55%	35%	32%	39%	23%	40%	0.38	0.00
Addition Facts	Timed	CPM	8.84	5.67	3.29	3.79	5.54	1.89	4.99	0.73	0.00
Addition Process	Untimed	%	50%	29%	12%	11%	38%	18%	31%	0.67	0.00
Subtraction Facts	Untimed	%	65%	45%	24%	25%	41%	20%	38%	0.55	0.00
Subtraction Facts	Timed	CPM	8.08	4.39	2.09	2.69	6.00	1.70	4.98	0.87	0.00
Subtraction Process	Untimed	%	40%	20%	7%	6%	33%	14%	27%	0.70	0.00
Word Problems	Untimed	%	47%	27%	14%	12%	33%	15%	29%	0.62	0.00
Shape Recognition	Untimed	%	14%	5%	2%	3%	12%	1%	15%	0.74	0.00

6.9 Score Movement Analysis

6.9.1 Baseline Performance Distribution - Count of Students

Table 16: Baseline Performance Category - Count of Students

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Number Comparison (untimed)	175	66	233	243	368	159	46	190	243	342
Counting in Bundles (untimed)	910	0	87	55	33	760	0	102	46	72
Missing Number (untimed)	512	120	264	127	62	423	134	223	119	81
Addition facts (untimed)	505	0	151	128	301	442	0	132	128	278
Addition facts (timed)	344	0	132	162	447	293	0	102	151	434

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Addition Process (untimed)	693	125	131	129	7	665	116	90	97	12
Subtraction facts (untimed)	550	0	273	95	167	489	0	240	104	147
Subtraction facts (timed)	512	1	131	101	340	451	0	95	85	349
Subtraction Process (untimed)	869	96	57	53	10	807	74	43	48	8
Word problem (untimed)	549	324	126	65	21	529	272	99	60	20
Shape Recognition (untimed)	1018	0	62	5	0	875	0	99	6	0

6.9.2 Endline Performance Distribution - Count of Students

Table 17: Endline Performance Category - Count of Students

Sub-task	Intervention					Comparison				
	L0	L1	L2	L3	L4	L0	L1	L2	L3	L4
Number Comparison (untimed)	125	9	55	150	746	168	26	91	190	505
Counting in Bundles (untimed)	288	0	159	135	503	445	0	139	109	287
Missing Number (untimed)	79	57	242	246	461	215	95	272	182	216
Addition facts (untimed)	100	0	81	133	771	233	0	110	153	484
Addition facts (timed)	60	1	33	79	912	165	1	50	114	650
Addition Process (untimed)	138	113	181	327	326	348	141	161	249	81
Subtraction facts (untimed)	110	0	151	155	669	275	0	202	145	358
Subtraction facts (timed)	128	0	34	45	878	273	2	59	60	586
Subtraction Process (untimed)	264	142	146	330	203	478	190	124	149	39
Word problem (untimed)	160	204	143	310	268	349	202	142	203	84
Shape Recognition (untimed)	730	0	268	87	0	845	0	126	9	0

6.9.3 Student Score Movement from Baseline to Endline - Intervention

Table 18: Score Movements - Intervention

>10% Downward Movement   >10% Upward Movement

Task	Baseline Performance Category	Endline Performance Category				
		L0	L1	L2	L3	L4
Number Comparison (untimed)	L0	13%	1%	3%	16%	68%
	L1	21%	0%	11%	14%	55%
	L2	12%	2%	7%	19%	60%
	L3	10%	1%	5%	14%	71%
	L4	10%	1%	4%	10%	76%
Counting in Bundles (untimed)	L0	29%	0%	15%	13%	43%
	L1	-	-	-	-	-
	L2	16%	0%	14%	16%	54%
	L3	16%	0%	5%	9%	69%
	L4	15%	0%	9%	6%	70%
Missing Number (untimed)	L0	10%	6%	23%	22%	39%
	L1	7%	7%	21%	18%	48%
	L2	5%	3%	22%	23%	47%
	L3	6%	6%	27%	28%	34%
	L4	2%	3%	13%	19%	63%
Addition facts (untimed)	L0	12%	0%	11%	13%	64%
	L1	-	-	-	-	-
	L2	9%	0%	6%	17%	69%
	L3	4%	0%	6%	16%	74%
	L4	7%	0%	3%	7%	83%
Addition facts (timed)	L0	6%	0%	3%	8%	83%
	L1	-	-	-	-	-
	L2	8%	0%	5%	14%	73%
	L3	6%	0%	5%	6%	83%
	L4	5%	0%	2%	5%	88%

Task	Baseline Performance Category	Endline Performance Category				
		L0	L1	L2	L3	L4
Addition Process (untimed)	L0	14%	12%	17%	29%	28%
	L1	11%	10%	22%	30%	26%
	L2	12%	5%	14%	34%	36%
	L3	6%	7%	14%	33%	40%
	L4	14%	29%	14%	14%	29%
Subtraction facts (untimed)	L0	12%	0%	15%	14%	59%
	L1	-	-	-	-	-
	L2	9%	0%	15%	16%	60%
	L3	5%	0%	12%	15%	68%
	L4	8%	0%	11%	13%	68%
Subtraction facts (timed)	L0	13%	0%	3%	4%	80%
	L1	0%	0%	0%	0%	100%
	L2	9%	0%	4%	6%	81%
	L3	8%	0%	4%	4%	84%
	L4	13%	0%	3%	3%	81%
Subtraction Process (untimed)	L0	25%	14%	14%	30%	17%
	L1	21%	11%	16%	31%	21%
	L2	26%	9%	12%	33%	19%
	L3	19%	9%	11%	32%	28%
	L4	20%	0%	0%	30%	50%
Word Problems (untimed)	L0	16%	20%	13%	30%	22%
	L1	15%	19%	14%	26%	25%
	L2	12%	14%	17%	28%	29%
	L3	9%	15%	9%	34%	32%
	L4	5%	24%	5%	24%	43%
Shape Recognition (untimed)	L0	67%	0%	24%	9%	0%
	L1	-	-	-	-	-
	L2	64%	0%	29%	7%	0%
	L3	20%	0%	60%	20%	0%
	L4	-	-	-	-	-

6.9.4 Student Score Movement from Baseline to Endline - Comparison

Table 19: Score Movements – Comparison

>10% Downward Movement   >10% Upward Movement

Task	Baseline Performance Category	Endline Performance Category				
		L0	L1	L2	L3	L4
Number Comparison (untimed)	L0	16%	3%	9%	23%	50%
	L1	13%	2%	9%	39%	37%
	L2	17%	4%	15%	21%	43%
	L3	21%	3%	7%	20%	49%
	L4	15%	2%	8%	14%	61%
Counting in Bundles (untimed)	L0	48%	0%	13%	11%	28%
	L1	-	-	-	-	-
	L2	37%	0%	15%	9%	39%
	L3	37%	0%	20%	17%	26%
	L4	40%	0%	18%	14%	28%
Missing Number (untimed)	L0	27%	10%	29%	17%	18%
	L1	20%	10%	23%	18%	28%
	L2	18%	12%	29%	21%	21%
	L3	16%	8%	25%	23%	29%
	L4	19%	6%	31%	16%	28%
Addition facts (untimed)	L0	27%	0%	11%	17%	45%
	L1	-	-	-	-	-
	L2	30%	0%	9%	11%	50%
	L3	26%	0%	13%	13%	48%
	L4	15%	0%	11%	17%	57%
Addition facts (timed)	L0	23%	0%	6%	12%	59%
	L1	-	-	-	-	-
	L2	13%	1%	7%	15%	65%
	L3	21%	0%	4%	11%	64%
	L4	12%	0%	4%	11%	73%
Addition Process (untimed)	L0	38%	15%	16%	24%	7%
	L1	28%	18%	19%	27%	9%
	L2	28%	12%	17%	32%	11%
	L3	34%	9%	16%	31%	9%
	L4	25%	17%	25%	0%	33%

Task	Baseline Performance Category	Endline Performance Category				
		L0	L1	L2	L3	L4
Subtraction facts (untimed)	L0	34%	0%	21%	13%	31%
	L1	-	-	-	-	-
	L2	26%	0%	18%	15%	41%
	L3	17%	0%	24%	17%	41%
	L4	20%	0%	19%	18%	42%
Subtraction facts (timed)	L0	30%	0%	7%	7%	55%
	L1	-	-	-	-	-
	L2	29%	0%	1%	6%	63%
	L3	33%	0%	6%	7%	54%
	L4	23%	0%	6%	5%	66%
Subtraction Process (untimed)	L0	51%	20%	12%	14%	4%
	L1	43%	23%	14%	15%	5%
	L2	33%	12%	19%	33%	5%
	L3	31%	21%	19%	25%	4%
	L4	50%	0%	25%	25%	0%
Word Problems (untimed)	L0	40%	19%	13%	20%	8%
	L1	35%	24%	15%	17%	10%
	L2	24%	19%	22%	28%	6%
	L3	32%	22%	8%	28%	10%
	L4	10%	20%	30%	25%	15%
Shape Recognition (untimed)	L0	87%	0%	12%	1%	0%
	L1	-	-	-	-	-
	L2	78%	0%	22%	0%	0%
	L3	100%	0%	0%	0%	0%
	L4	-	-	-	-	-

6.10 Analysis of Literacy Investigation

6.10.1 Overview

The literacy assessment was administered alongside the numeracy assessment and covered an identical sample of students. The final analytical sample includes only those students who participated in both literacy and numeracy assessments across the Baseline and Endline rounds. Literacy outcomes were measured using a contextualised version of the Early Grade Reading Assessment (EGRA), which was administered uniformly to students in Grades 2 and 3 and remained consistent across both rounds. To ensure alignment with the programme context, the assessment tool was adapted and contextualised and translated into Hindi.

This literacy investigation is indeterminate and should be interpreted with appropriate caution. All the literacy sub-tasks relied on audio-based assessments. Variations in audio clarity and inconsistencies in recording conditions have influenced the measurement precision.

6.10.2 Indicative Investigation

The Mindspark literacy assessment indicates positive gains across most assessed domains, with largest gains in comprehension tasks i.e. Reading Comprehension (untimed: 0.38 SD) and Listening Comprehension (0.27 SD). Decoding skills improved as well, with statistically significant gains in Letter Recognition (untimed: 0.28 SD), reflecting improvements in accuracy, while gains in timed Letter Recognition were smaller (0.10 SD) and statistically insignificant. Familiar Word Reading demonstrated similar gains across both untimed (0.15 SD) and timed (0.14 SD) formats. Oral Vocabulary also showed statistically significant improvement (0.17 SD).

Table 20: Mindspark Literacy Results

Mindspark - Literacy											
Task	Task Type	Unit	Endline Avg.		Baseline Avg.		Delta (EL - BL)		Pooled SD	Effect Size (SD)	p-value[50]
			I	C	I	C	I	C			
Listening Comprehension	Untimed	%	48%	28%	32%	21%	16%	6%	35%	0.27	0.00
Oral Vocabulary	Untimed	%	71%	68%	62%	64%	9%	4%	25%	0.17	0.00
Letter Recognition	Untimed	%	64%	59%	38%	43%	25%	16%	35%	0.28	0.00
Letter Recognition	Timed	CPM	22.8	22.5	15.1	16.1	7.7	6.3	14.28	0.10	0.06
Familiar Word Reading	Untimed	%	54%	51%	28%	31%	26%	21%	38%	0.15	0.00
Familiar Word Reading	Timed	CPM	14.3	13.8	6.8	7.9	7.5	5.9	11.26	0.14	0.01
Nonword Reading	Timed	CPM	13.0	13.1	5.9	6.9	7.1	6.2	11.19	0.08	0.11
Oral Reading Fluency	Timed	CPM	26.0	24.4	7.8	9.5	18.2	14.9	32.69	0.10	0.06
Reading Comprehension	Untimed	%	36%	26%	9%	10%	27%	17%	28%	0.38	0.00

In contrast, gains in both Nonword Reading (0.08 SD) and Oral Reading Fluency (0.10 SD) were not statistically significant. Overall, the results indicate that Mindspark was effective in strengthening comprehension and recognition skills, with limited impacts on fluency and decoding of unfamiliar words.

50 A p-value of less than 0.05 is considered statistically significant.

LiftEd Transforming Foundational Literacy and Numeracy

EDTECH ACCELERATOR

To learn more about the

LiftEd EdTech Accelerator,

visit our website:

www.edtechaccelerator.org

Contact us at:
edtechaccelerator@centralsuarefoundation.org